



申请代码	F020709
接收部门	
收件日期	
接收编号	

国家自然科学基金 申 请 书

(2019 版)

资助类别：	面上项目		
亚类说明：			
附注说明：			
项目名称：	群智数据融合网络关键技术研究		
申 请 人：	甘小莺	电 话：	021-34205018
依托单位：	上海交通大学		
通讯地址：	上海市东川路800号电信楼1-437室		
邮政编码：	200240	单位电话：	021-34206809-188
电子邮箱：	ganxiaoying@sjtu.edu.cn		
申报日期：	2019年02月03日		

国家自然科学基金委员会



基本信息

申请人信息	姓名	甘小莺	性别	女	出生年月	1978年07月	民族	汉族
	学位	博士	职称	副教授	每年工作时间（月）		8	
	是否在站博士后	否		电子邮箱	ganxiaoying@sjtu.edu.cn			
	电话	021-34205018		国别或地区	中国			
	个人通讯地址	上海市东川路800号电信楼1-437室						
	工作单位	上海交通大学/电子信息与电气工程学院						
	主要研究领域							
依托单位信息	名称	上海交通大学						
	联系人	章俊梅	电子邮箱	amyzhang@sjtu.edu.cn				
	电话	021-34206809-188	网站地址	www.sjtu.edu.cn				
合作研究单位信息	单位名称							
项目基本信息	项目名称	群智数据融合网络关键技术研究						
	英文名称	Reaearch on key technology of data enabled crowdsourcing networks						
	资助类别	面上项目				亚类说明		
	附注说明							
	申请代码	F020709. 新型感知计算及网络						
	基地类别							
	研究期限	2020年01月01日 -- 2023年12月31日				研究方向：群智感知网络		
	申请直接费用	81.3000万元						
中文关键词		群智计算；感知智能化；数据推断；网络嵌入；物联网						
英文关键词		Crowdcomputing; intelligent computing; data infusion; network embedding; IoT						



中文摘要	<p>在群智数据融合网络中，网络嵌入技术被引入到群智系统，对于提高数据计算效率，提升任务管理能力，实现用户-群智平台的合作共赢具有重要价值。一方面，可以促使群智场景和异构数据的高效融合；另一方面，可以提升用户优选及资源分配能力。然而，如何有效表征复杂群智场景中的实体间关联，基于实体嵌入表征实现数据真实值精准计算，利用实体表征促进任务高效分发及资源优化配置等问题都亟待解决。为解决上述问题，拟实现以下目标：</p> <ol style="list-style-type: none">1) 设计基于群智场景元路径的网络嵌入算法，实现实体属性和结构的数据化表征，跨越高维结构到数据空间的映射障碍，探索物理感知和数据认知的融合；2) 设计融合数据与实体相关性的推测算法，通过探索物理空间的时空属性对数据空间的相关性影响，提高真实值推测的准确度，降低决策制定中信息不确定带来的损失；3) 提出基于实体多属性的在线任务分配机制，增强群智网络区域管理能力，实现资源的有效利用和任务的高效分发。
英文摘要	<p>A data enabled crowdsourcing network is defined as a crowdsourcing system combined with network embedding technology. Such network could achieve user-platform win-win cooperation through both the improved task management ability and higher computational efficiency. On one hand, the network embedding can promote the highly efficient integration of crowdsourcing scenes and heterogeneous data; on the other hand, it also enhances the user selection and resource allocation abilities in crowdsourcing network. However, the following problems still remain unsolved: How to effectively characterize entity relationships among complex crowdsourcing scenes? How to accurately calculate the truth value of data based on the entity data embedding representation? How to achieve efficient task allocation and resource distribution. In order to solve these problems, this project aims to achieve the following objectives:</p> <ol style="list-style-type: none">1) Propose network embedding algorithm based on meta path of crowdsourcing scenes, which could achieve the data based representation of both entity attributes and structures. Further, we could across the mapping barriers from high dimension structures to data space, and explore the fusions of physical perception and data recognition.2) Design inference algorithms fusing data and entity correlations. By exploring the influences from physical space to data space brought by the temporal-spatial attributes, we aim to improve the truth inference accuracy and reduce the loss caused by uncertain information in decision-making process.3) Propose entity multi-attribute based online task allocation mechanism, so as to enhance the crowdsourcing network area management ability and achieve highly efficient resource utilization and task distribution.



项目组主要参与者（注：项目组主要参与者不包括项目申请人）

编号	姓名	出生年月	性别	职 称	学 位	单位名称	电话	电子邮箱	证件号码	每年工作 时间（月）
1	张伟楠	1989-02-25	男	讲师	博士	上海交通大学	34205018	wnzhang@sjtu.edu.cn	5*****7	6
2	范桂云	1993-07-26	女	博士生	学士	上海交通大学	4205018	fgy726@sjtu.edu.cn	3*****4	10
3	秦志达	1991-06-27	男	博士生	学士	上海交通大学	34205018	zanderqin@163.com	1*****3	10
4	王雄	1991-12-26	男	博士生	学士	上海交通大学	4205018	wangxionsjtu@sjtu.edu.cn	4*****X	10
5	李雨晴	1991-05-11	女	博士生	学士	上海交通大学	34205018	liyuqing@sjtu.edu.cn	3*****6	10
6	王彪	1990-02-08	男	博士生	硕士	上海交通大学	34205018	sweetmvp@sjtu.edu.cn	3*****9	9

总人数	高级	中级	初级	博士后	博士生	硕士生
7	1	1			5	



国家自然科学基金项目资金预算表（定额补助）

项目申请号：

项目负责人：甘小莺

金额单位：万元

序号	科目名称	金额
	(1)	(2)
1	项目直接费用合计	81.3000
2	1、设备费	7.2000
3	(1)设备购置费	7.20
4	(2)设备试制费	0.0000
5	(3)设备升级改造与租赁费	0.0000
6	2、材料费	5.30
7	3、测试化验加工费	4.00
8	4、燃料动力费	4.80
9	5、差旅/会议/国际合作与交流费	18.80
10	6、出版/文献/信息传播/知识产权事务费	9.40
11	7、劳务费	28.80
12	8、专家咨询费	3.00
13	9、其他支出	0.0000



预算说明书（定额补助）

（请按《国家自然科学基金项目资金预算表编制说明》中的要求，对各项支出的主要用途和测算理由及合作研究外拨资金、单价 ≥ 10 万元的设备费等内容进行详细说明，可根据需要另加附页。）

1. 设备费 共计7.2万元

用于建立基于生成式对抗的深度学习群智系统，并对其隐私保护算法进行仿真计算，需购置笔记本2台，每台约0.8万；工作站1台，约4万元；台式工作站2台，每台约0.8万。

2. 材料费共计5.3万元

实验仿真需购买移动终端10台，每台0.3万；GPU板卡2块，每块0.9万，共计1.8万元；硒鼓8个，每个500元，共计0.4万元；打印纸10箱，每箱100元，共计0.1万元。

3. 测试化验加工费共计4万元

生成式对抗网络面向隐私异质性复杂环境下的集中测试费用，其中深度学习参数性能测试约2万元；大用户量异质模型的激励机制性能测试约1万元；包含500个社交网络用户的隐私泄露评估约1万元。

4. 燃料动力费共计4.8万元

利用多核信号处理平台模拟用户群智标记、隐私关联等复杂网络行为，利用GPU工作站产生大量虚拟的连接和信息以达到模拟用户信息收集行为的目的，对群智系统隐私保护进行协议与性能验证。多核信号处理平台和GPU工作站功耗约2600瓦，每月用电量约1872度，按照0.641元/度计算，并考虑每年工作10个月，电费用量每年约1.2万元。

5. 差旅/会议/国际合作与交流费共计18.8万元

1) 差旅费：课题组成员赴北京，苏州，杭州，西安，郑州，沈阳等国内城市参加研讨会和学术活动所产生的差旅费用，每人每次约2500元，每年10人次，约10万元。

2) 会议费：课题组计划4年内组织开展2次小型学术研讨会，按照国家公布的《中央和国家机关会议费管理办法》，邀请国内核心专家与学者参加：0.04万元/人/天/次 $\times 10$ 人 $\times 1$ 天 $\times 2$ 次=0.8万元。

3) 国际合作与交流费：发表高水平国际会议论文需要现场宣读论文，计划4年进行4人次国际学术交流，按照《因公临时出国（境）经费管理办法》，每次约1.5万元，4年共计6万元。邀请境外专家来做指导咨询的国际机票费、住宿费等，平均每2年1人次，每人次平均1万元，4年共计2万元。

6. 出版/文献/信息传播/知识产权事务费共计9.4万元

版面费，每篇国际会议/期刊论文约900美元，10-15篇约6-9万元；专利申请维护费每项约0.6万元，4项约2.4万元、资料费、网络费等4年1万元。

7. 劳务费共计28.8万元

用于发放课题组研究生的劳务费，包括5名博士生（共36个月），及1名硕士生（共36个月），按照所在单位发放标准，测算如下：博士生0.15万元/人/月 $\times 5$ 人 $\times 36$ 月+硕士生0.05万元/人/月 $\times 1$ 人 $\times 36$ 月=28.8万元。

8. 专家咨询费共计3万元

根据《中央财政科研项目专家咨询费管理办法》，支付给临时聘请的咨询专家的费用：0.15万元/人/天/次 $\times 5$ 人 $\times 1$ 天 $\times 4$ 次=3万元。



报告正文

(一) 立项依据与研究内容 (建议 8000 字以下):

1. 项目的立项依据 (研究意义、国内外研究现状及发展动态分析,需结合科学研究发展趋势来论述科学意义;或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录);

1.1 研究意义和必要性

在移动物联网以及大数据时代,群智网络在公共安全、城市管理、商务智能、环境监测、智能交通等领域发挥了重要的作用。随着嵌入式设备、无线传感网络、智能移动终端等的飞速发展,集成感知、计算和通信能力的移动群智系统广受关注。在此背景下,面向城市和社会等物理空间的群智数据融合网络成为当前信息领域的研究热点。然而,由于其应用场景极为丰富,群智数据融合网络呈现出**终端异构性、实体交互性、互联多样性**等显著特征[1],现有的模型和方法难以满足群智系统海量数据计算和异构任务管理的要求,因此亟需探索新的计算模型和方法。网络嵌入作为一种新兴的数据信息表征机制,在信息技术领域崭露头角。例如,亚马逊和爱彼迎(Airbnb)分别将网络嵌入应用到商品点评和房源排序系统中进行评价质量评估和个性化推荐[2-3]。

将网络嵌入技术引入到群智系统中,对于提高数据计算效率,提升任务管理能力,实现用户-群智平台的合作共赢具有重要的价值。在复杂群智系统中,网络嵌入技术通过将多维实体属性嵌入低维向量空间,可以实现对抽象实体关联的具象表征,拓展群智系统应用场景。而群智系统与网络嵌入相结合,一方面促使群智场景和异构数据的高效融合,挖掘数据潜在时空特征,促进复杂数据的高效计算与预测;另一方面基于获取的数据空间信息与网络实体表征,在有限预算的情况,基于多样化任务需求,提升用户优选及资源分配管理能力。考虑到网络规模的急剧增长和演进特性,数据计算复杂度过高以及噪声问题,提升群智网络的服务质量仍然有很多问题亟待解决:**如何有效地表征复杂群智场景中的实体间关联?如何基于实体嵌入表征挖掘数据关联特征,并实现数据真值精准推断与计算?如何利用获取的数据信息及实体表征促进任务高效分发以及资源优化配置?**

针对上述待解决的问题,本项目拟面向群智数据融合网络开展系统性与前瞻性研究,其



研究意义在于：

- 1) 在复杂群智网络下，通过对多样化群智场景的语义解析，实现实体属性和结构的数据化表征，跨越高维结构到数据空间的映射障碍，探索物理感知和数据认知的融合。
- 2) 融合相关性特征的真值推断，通过探索物理空间的时空属性对数据空间的相关性影响，结合实体层面的用户互联，提高真值推断的准确度，显著降低决策制定中信息不确定带来的损失。
- 3) 基于真值推断信息，设计异构时空任务分配和资源配置策略，利用实体相关性进一步降低数据冗余度、优化数据质量，这对于实现资源的有效利用、任务的高效分发具有重要的研究意义。

1.2 技术发展背景和现状

作为近年来国内外学术界研究的热点，群智网络在现有研究中，一般基于单一场景，假设用户关系独立，数据模型较单一。不同于以往的研究场景，群智数据融合网络具有终端异构性、实体交互性、互联多样性等特征，复杂的应用场景又对群智系统数据计算和任务管理能力提出了更高的要求，然而，网络嵌入技术的兴起为解决这一难题提供了全新的思路。本项目主要目的在于为基于网络嵌入的群智数据融合网络提供一套理论框架和分析方法，在网络嵌入、真值推断、群智任务分配方面的国内外主要研究成果如下：

1.2.1 国外相关研究

(1) 网络嵌入研究

单一时间窗下的异构网络嵌入问题。在传统的网络分析中，因为网络高维和结构化的特点，研究人员只能使用度分布，独立级联模型这种结构层次的特征或算法来挖掘网络中的隐藏信息。这种方法，一方面只能反应连接关系下的结构特征，而没有考虑到节点的属性特征；另一方面算法复杂度高，无法适应大规模的网络处理任务。近年来，针对这些问题，网络嵌入逐渐成为了人们关注的热点。Hu 等学者[4]针对以往模型很少学习推荐任务中路径或元路径的显式表示，以及未考虑元路径和交互中涉及的用户-项对之间的相互影响问题，提出了一种具有共同关注机制的新型深度神经网络，能够利用丰富的基于元路径的上下文进行 top-N 推荐。Cui 等学者基于奇异值分解理论，设计了一种名叫 HOPE[5]的网络嵌入算法，该方法利用特征向量提取，将高维的网络结构降维到低维的向量空间中，并且由于算法具有可伸缩性，所以可以保留网络结构的高阶近似以及捕捉网络的非对称传递性。Bryan Perozzi 等学



者,提出了一种学习网络中顶点的潜在表示的新方法 DeepWalk [6]。这些潜在的特征,通过在一个连续的向量空间中编码社交关系得到,可以很容易被统计模型利用。Dong 等学者研究了异构网络中的表征学习问题,提出了一种基于元路径的网络嵌入算法 metapath2vec [7],该模型将基于元路径的随机游走形式化,构建节点的异构邻域,然后利用异构 skip-gram 模型执行节点嵌入,该方法具有计算复杂度低,能够抓取语义信息等优点。

基于动态演进下的网络嵌入问题。传统的网络嵌入算法,往往只关注固定的网络结构。但实际上,网络具有自我演进的特性,网络中的节点和连接根据时间的变化,会呈现不同的交互特性。所以静态场景下的网络嵌入无法反应出这种演进的过程,如何通过分析动态网络,来挖掘时变场景下的潜在价值,就成了一个关键问题。Li 等学者根据网络的时变特性,提出了一种动态分布的嵌入框架 DANE [8],该方法根据用户过去一段时间的特征变化,在线更新用户当前的嵌入结果,所以考虑到了时变中的信息。Zhou 等学者提出了一种 DynamicTriad 算法[9],该方法根据网络中三元组从开环到闭环的变化信息,使模型能够捕捉网络动态,学习不同时间步长的每个顶点的表示向量。

(2) 基于相关特征的真值推断

感知任务敏感的异质用户能力建模问题。刻画用户质量是真值推断的基础,最直观的建模方法是利用概率参数表示用户数据的正确概率,概率参数在 $[0, 1]$ 区间,值越大表示用户可靠性越高。Rodrigues 等学者设计了混淆矩阵,更加精细地刻画用户质量[10],假设任务的真值取自 1 个答案集合,混淆矩阵的第 j 行表示当真值为第 j 个答案时,用户数据在 1 个答案集合上的概率分布。用户权重也是衡量用户质量的重要指标,当用户质量更高时,其感知数据在真值推断中的权重更大[11]。数值型任务推断中,偏差和方差是衡量用户质量的重要指标[12],偏差代表用户数据相对真值的平均误差,方差代表偏差的波动。CHENG 等学者考虑多类型任务,并区分用户对不同类型任务的技能(能力),通过技能向量精细刻画用户质量[13]。

融合相关性的真值推断问题。随着社交网络的兴起,实体间的相关联系越来越紧密,特别是位置服务的爆炸性增长,基于地理位置的群智系统近年来也受到了学界的关注。Jiang 等学者通过利用时空相关特征,将用户组间成为社区,从而实现应用社区分布式感知环境数据的目的,由于社区的划分以及真值推断都考虑了数据的时空属性,从而真值推断的准确性能大幅提高[14]。Wang 等学者考虑实体间的社交关系,分析用户数据与社交关系的依赖性,通过建立最大似然概率模型,结合社交相关精准推断任务真值[15]。考虑到参与感知的设备数量巨大,任务真值随时间动态变化,Zhang 等学者研究了动态真值推断问题,其中用



户分布式收集数据,并集合用户间的社交关系,利用隐马尔可夫链增量更新推断结果,从而实现在线分布式推断的目的 [16]。

(3) 基于实体属性的任务分配机制研究

考虑实体相关性的任务分配问题。群智系统中多维异构实体之间往往存在着高度的相互关联性,严重降低了传统单属性任务分配模型泛化能力。基于实体相关性设计任务分配策略可以有效减少冗余信息造成的误差,提升数据质量的准确度。He 等学者针对城市信号地图的构建问题提出了基于联合贝叶斯压缩感知的群智激励机制[17],通过挖掘不同测量点在空间、信号、时间维度的相关性,实现信号质量和成本间的均衡,保证激励算法对信号图环境动态变化的稳健性。Zhang 等学者针对感测变量的多个属性间的依赖性和约束问题进行研究,利用在线学习和贝叶斯推断理论,设计任务优先级估计和多属性优先级整合策略,得到了基于多属性相关性的最优任务分配机制[18]。Wang 等学者提出了稀疏任务分配框架 SPACE-TA [19],利用同类型和不同类型感测数据之间的相关性来减少感知任务分配的数量,解决了如何在稀疏感知区域情况下实现高质量感知数据的难题。考虑到用户在决策时往往表现出显著的多样个性化,传统基于用户完全理性的决策先验模型不再适应, Karaliopoulos 等学者首先建模任务的可控属性和非可控属性特征,利用机器学习技术挖掘用户个人偏好信息,预测用户对异构多属性的任务执行概率,优化任务和报酬分配方案,最大化用户执行任务的概率[20]。

基于时空相关性的任务分配问题。移动群智系统需要依托大量参与用户共同完成众多群智任务,而充分挖掘和利用参与用户间地理位置和群智任务的时空相关性能够帮助实现用户和任务之间的高效准确匹配,从而极大的提升群智系统的服务性能。随着位置服务的兴起,基于地理位置的群智系统近年来也越来越多的受到了学界的关注,基于时空相关性的任务分配问题已经成为了近年来国内外研究的热点。Wang 等学者研究了用户位置已知和任务动态到达下的时空任务分配策略[21],研究了线下和线上两种场景下的公平分配问题,结合李雅普诺夫优化实现了系统稳定下的效用最大化。Zhu 等学者利用群智用户的动态移动属性和感知能力的异质性[22],考虑了任务的时空因素,设计了中心式任务在线规划调度算法,该算法结合了用户当前时刻的状态以及历史感知数据的信息,满足了混合城市感知的不同需求。Wang 等学者针对任务的特性以及感知情景的不同[23],利用感知任务需求的异质性和用户动态移动属性,建模并存多任务的时空相关性分析模型,提出了两级任务分配策略,通过在相似任务和无相似任务间的分级搜索实现了最优任务分配以最大化收集数据的质量。此外, Wang 等学者也设计了先分解后合并的搜索框架和合理的激励机制,以减小搜索空间



实现计算的高效和最小化激励开销。

1.2.2 国内的相关研究

近年来,国内学者也开始关注网络嵌入等方面的研究。Tu [28]等学者针对现有网络嵌入模型仅能应用于双实体关系的嵌入的缺陷,提出了深度超网络嵌入模型,解决了基于多实体关系的超网络嵌入问题,进一步拓展了网络嵌入的实际应用场景。Zhang [27]等学者针对网络嵌入中高阶邻近关系难以保存的问题,提出了 AROPE 网络嵌入模型,实现了对于网络中任意阶邻接关系的高效嵌入,并从理论上证明了该方法在准确性和复杂度上均实现了最优解。群智感知系统近年来在国内也得到了广泛的研究。在文献[25]中,Chai 等学者针对群智系统中任务所面临的多样化数值需求,提出了基于激励的实体选择机制,在对实体可靠性进行数据准确度和冗余性双重验证,实现了任务与实体高效匹配,提高了群智系统的性能。Zhao [29]等学者针对群智系统中在线用户激励需求,将有限预算情况下在线用户优选问题建模为在线拍卖问题,并且提出了 OMZ 和 OMG 两种拍卖机制,从理论上实现了在线用户优选的计算高效性,预算可行性,个体合理性以及数据可靠性等要求。考虑到任务多属性对于实体选择的影响,任务附属属性越复杂,实体选择要求也越高,复杂度越大,文献[24]研究了基于任务属性关联度的优选算法,在实现高效实体选择的同时,降低了处理复杂度。用户原始数往往包含噪声,难以满足系统需求,Yang[26]等学者将数据真实值推测和盈余分享相结合,采用无监督学习建模用户数据质量模型,筛选高质量数据,用采用合作博弈机制实现高效盈余分享,达到了高质量和高盈余的双赢效果。

1.2.3 小结

上述工作涵盖了网络嵌入和群智感知系统研究的诸多方面,给予我们很大的启发,为我们进一步研究提供了理论基础和新思路。然而,本课题与上述工作不同之处在于,我们拟用网络嵌入、深度学习、贝叶斯网络、优化理论等工具,对群智数据融合网络进行系统性地分析和性能优化,利用网络嵌入模型实现群智场景实体关系表征,进一步融合数据空间相关性实现用户真值精准推断,以及基于实体多属性的任务高效分配。

参考文献

- [1] 郭斌,翟书颖,於志文,周兴社,“群智大数据:感知、优选与理解”,大数据,vol. 5, pp, 60-72, 2017.
- [2] M. Grbovic, H. Cheng, “Real-time Personalization using Embeddings for Search Ranking at



- [Airbnb](#)”, in Proc. ACM SIGKDD, Aug. 2018.
- [3] K. Parisa, J. Caverlee, and A. Squicciarini, “[Combating Crowdsourced Review Manipulators: A Neighborhood-Based Approach](#)”, in Proc. ACM WSDM, Feb. 2017.
- [4] Hu. B, Shi. C, Zhao W. X, and Yu. P. S, “[Leveraging meta-path based context for top-n recommendation with a neural co-attention model](#)”. In Proc. ACM SIGKDD, Jul. 2018.
- [5] Ou. M, Cui. P, Pei. J, Zhang. Z, and Zhu. W, “[Asymmetric transitivity preserving graph embedding](#)”, in Proc. ACM SIGKDD, Aug. 2016.
- [6] Perozzi. B, Al-Rfou. R, and Skiena. S, “[Deepwalk: Online learning of social representations](#)”, In Proc. ACM SIGKDD, Aug. 2016.
- [7] Dong. Y, Chawla. N. V, and Swami. A, “[metapath2vec: Scalable representation learning for heterogeneous networks](#)”, In Proc. ACM SIGKDD, Aug. 2017.
- [8] Li. J, Dani. H, Hu. X, Tang. J, Chang. Y, and Liu. H, “[Attributed network embedding for learning in a dynamic environment](#)”, In Proc. ACM CIKM. Nov. 2017.
- [9] Zhou. L, Yang. Y, Ren. X, Wu. F, and Zhuang. Y, “[Dynamic Network Embedding by Modeling Triadic Closure Process](#)”, in Proc. ACM AAAI, Apr. 2018.
- [10] F. Rodrigues, and F. C. Pereira, “[Deep Learning from Crowds](#)”, in Proc. AAAI. Apr. 2018.
- [11] H. Jin, L. Su, and K. Nahrstedt, “[Theseus: Incentivizing truth discovery in mobile crowd sensing systems](#)”, in Proc. ACM MobiHoc, July 2017.
- [12] R. W. Ouyang, L. M. Kaplan, A. Toniolo, M. Srivastava, and T. J. Norman, “[Aggregating Crowdsourced Quantitative Claims: Additive and Multiplicative Models](#)”, IEEE Trans. on Knowledge and Data Engineering, vol. 28, no. 7, pp. 1621-1634, 2016.
- [13] P. Cheng, X. Lian, L. Chen, J. Han, and J. Zhao, “[Task Assignment on Multi-Skill Oriented Spatial Crowdsourcing](#)”, IEEE Trans. on Knowledge and Data Engineering, vol. 28, no. 8, pp. 2201-2215, 2016.
- [14] J. Bian, H. Xiong, Y. Fu, and S. K. Das, “[CSWA: Aggregation-Free Spatial-Temporal Community Sensing](#)”, in Proc. AAAI, Apr. 2018.
- [15] C. Huang, and D. Wang, “[Topic-aware social sensing with arbitrary source dependency graphs](#)”, in Proc. ACM IPSN, Apr. 2016.
- [16] D. Zhang, C. Zheng, D. Wang, D. Thain, C. Huang, X. Mu, and G. Madey, “[Towards Scalable and Dynamic Social Sensing Using A Distributed Computing Framework](#)”, in Proc. IEEE



ICDCS, July 2017.

- [17] S. He and K. G. Shin, “[Steering Crowdsourced Signal Map Construction via Bayesian Compressive Sensing](#)”, in Proc. IEEE INFOCOM, Apr. 2018.
- [18] Y. Zhang, D. Zhang, N. Vance, and D. Wang, “[Optimizing Online Task Allocation for Multi-Attribute Social Sensing](#)”, in Proc. ICCCN, July 2018.
- [19] L. Wang, D. Zhang, D. Yang, A. Pathak, C. Chen, X. Han, H. Xiong, and Y. Wang, “[SPACE-TA: Cost-Effective Task Allocation Exploiting Intradata and Interdata Correlations in Sparse Crowdsensing](#)”, ACM Trans. Intelligent Systems and Technology, vol.9, no.2, pp.1-28, Jan. 2018.
- [20] M. Karaliopoulos, I. Koutsopoulos, and M. Titsias, “[First Learn then Earn: Optimizing Mobile Crowdsensing Campaigns through Data-driven User Profiling](#)”, in Proc. ACM Mobihoc, July 2016.
- [21] X. Wang, R. Jia, X. Tian, X. Gan, “[Dynamic task assignment in crowdsensing with location awareness and location diversity](#)”, in Proc. IEEE INFOCOM, Apr. 2018.
- [22] Q. Zhu, M.Y.S., Uddin, N. Venkatasubramanian and C.H, Hsu, “[Spatiotemporal Scheduling for Crowd Augmented Urban Sensing](#)”, in Proc. IEEE INFOCOM, Apr. 2018.
- [23] L. Wang, Z. Yu, D. Zhang, B. Guo and C. H. Liu, “[Heterogeneous Multi-Task assignment in Mobile Crowdsensing Using Spatiotemporal Correlation](#),” in IEEE Trans. Mobile Computing, vol. 18, no. 1, pp. 84-97, Jan. 2019.
- [24] X. Weng, G. Li, H. Hu and J. Feng, “[Crowdsourced Selection on Multi-Attribute Data](#)”, in Proc. ACM CIKM, Nov. 2017.
- [25] C. Chai, J. Fan and G. Li. “[Incentive-Based Entity Collection using Crowdsourcing](#)”, in Proc. IEEE ICDE, May 2018.
- [26] S. Yang, F. Wu, S. Tang, X. Gao, B. Yang and G. Chen, “[On Designing Data Quality-Aware Truth Estimation and Surplus Sharing Method for Mobile Crowdsensing](#),” in IEEE Journal on Selected Areas in Communications, vol. 35, no. 4, pp. 832-847, Apr. 2017.
- [27] Z. Zhang, P. Cui, X. Wang, J. Pei, X. Yao and W. Zhu, “[Arbitrary-Order Proximity Preserved Network Embedding](#)”, In Proc. ACM KDD, Aug. 2018.
- [28] K. Tu, P. Cui, X. Wang, F. Wang, W. Zhu, “[Structural Deep Embedding for Hyper-Networks](#)”, in Proc. AAAI, Apr. 2018.



- [29] D. Zhao, X. Li, and H. Ma, “Budget-Feasible Online Incentive Mechanisms for Crowdsourcing Tasks Truthfully”, IEEE/ACM Tran. On Networking, vol. 24, no. 2, pp. 647-661, Apr. 2016.

2. 项目的研究内容、研究目标，以及拟解决的关键科学问题（此部分为重点阐述内容）；

2.1 研究内容

本项目研究内容如图 1 所示，我们运用网络嵌入、深度学习、贝叶斯网络和优化理论等工具，对群智数据融合网络展开前瞻性的研究。首先基于网络嵌入模型将高维物理空间映射到低维数据空间，通过嵌入无结构化数据挖掘实体间相关性；考虑数据在时间和空间维度的相关性，设计基于数据时空相关和实体社交互联的真值推断算法，融合多维度信息精确发现有效信息；基于真值推断信息与实体多维属性，提出实体优选策略，合理指导数据的收集，提高数据质量。三个内容相互促进，且系统回答了群智数据融合网络的三个重要问题。

群智数据融合网络

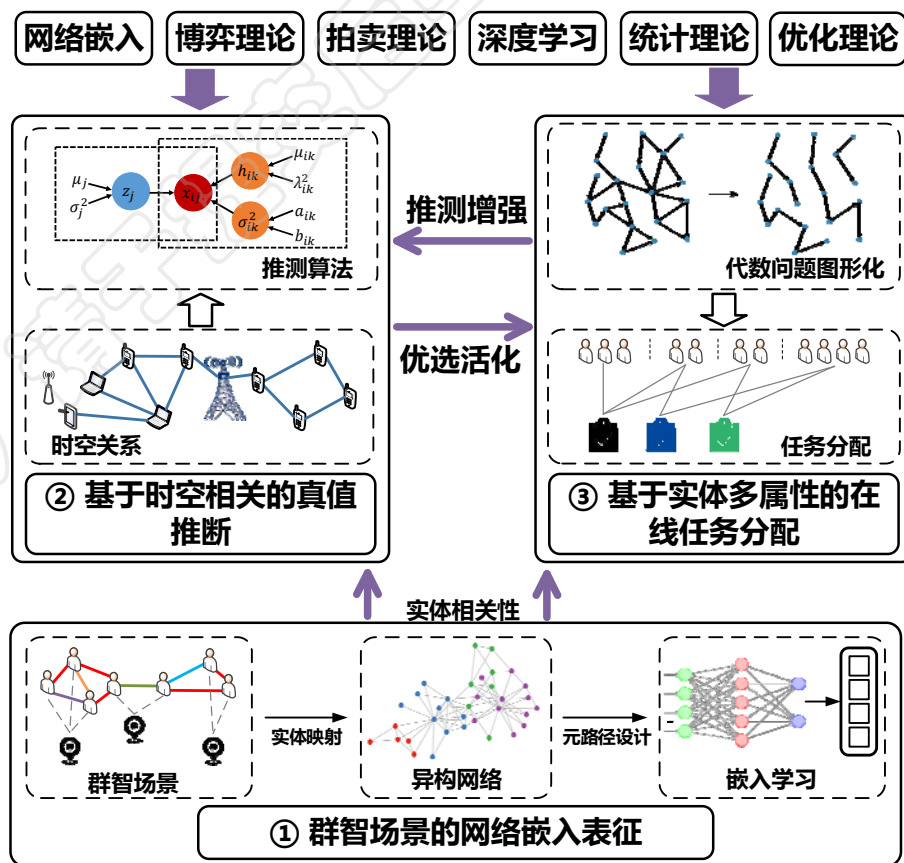


图 1. 研究内容框架图



2.1.1 群智场景的网络嵌入表征

在群智系统物理空间中,各种类型的实体,各种属性的联系,能够通过依赖关系相互组合,而形成一张包含丰富信息的网络。通过分析这个网络的结构和属性,能够获取到丰富的用户信息。但是群智网络通常具有高维、异构、演进等特征,特别是随着物联网和大数据时代的到来,这些特征使得网络的规模呈现指数型爆炸式增长,百万级乃至亿级的节点规模已经十分常见。超大规模网络结构使得经典的网络分析方法在此场景下的算法复杂度过高,无法有效的挖掘网络中有价值的信息。所以基于群智网络的异构性和大规模性,传统的研究方法难以体现其优势。

基于上述分析,本文提出一种基于元路径的异构网络嵌入研究方法来表征网络中的信息,包括以下三个方面:

- 1) 分析群智网络多维度特征数据,基于场景的先验知识,探索元路径设计方案,提升网络嵌入的效率。
- 2) 分析演进群智场景的嵌入表征,探究群智网络的发展趋势,设计有效的实体关系预测方法。
- 3) 基于元路径集合,研究随机游走过程中的交叉选择策略,提升嵌入结果承载不同维度信息的能力。

2.1.2 基于时空相关的真值推断

在群智网络中,用户收集数据通常包含有噪声,甚至相互冲突,真值推断能从用户数据中获取真实的信息,为进一步的决策奠定基础。用户感知数据涉及与物理环境的交互,此外用户间的相关性也影响数据收集过程,即用户物理空间的关联会映射成为数据空间的相关性。数据相关性主要体现在时间相关、空间相关,进而从多维度促进真值推断过程。用户作为数据感知的实体,其能力直接影响收集数据的质量,用户的异质性不仅作用于数据感知过程,而且关系到真值推断的设计。从具有噪声的用户数据中推断真值和挖掘有效信息需要合理的推断算法,一方面考虑数据质量的差异性,另一方面融合相关性特征,包括数据在物理空间的时空相关和用户在实体层面上的互联相关,总体上获得精确的推断结果。

基于上述分析,基于时空相关的真值推断研究包括以下三个方面:

- 1) 研究数据相关性的时空表征,从理论上描述时间相关、空间相关属性,阐述物理空间到数据空间相互作用的机理,分析数据相关性的计算逻辑。



- 2) 探索感知敏感的异质用户能力特征, 阐明用户能力和数据质量的关联特性, 根据感知任务类型建立用户能力模型, 解释用户能力对真值推断的影响。
- 3) 设计稳健的真值推断模型, 考虑用户能力异质性和数据质量差异性, 融合数据与实体相关性, 充分利用相关性提高推断准确度。

2.1.3 基于实体多属性的在线任务分配机制研究

移动群智网络通过招募大量用户来协同完成众多复杂任务。作为实体, 用户的数量及其质量直接影响着群智服务性能。因此如何选取实体进行任务分配成为一个亟待解决的关键问题。传统的任务分配研究大多是基于尽力而为服务, 不承载实体具体的属性特征或是面向单属性(如用户位置信息)来进行任务分配, 尚未形成全面智能的任务分配机制, 导致数据质量差、资源利用率低。目前许多群智计算应用, 其任务完成往往需要同时收集实体的多个属性。基于实体多属性的在线任务分配机制, 旨在通过挖掘实体属性的异构多态、复杂关联、动态涌现特征, 促进任务合理分配和实体优化选取, 实现高效、智能的群智网络。一方面, 异构多维实体之间存在显著的相关性, 导致模型的泛化能力弱。为避免次优解的出现, 亟需设计基于实体相关性的在线任务分配策略减少冗余信息造成的误差, 提升数据质量的精度; 此外, 不同实体属性对数据质量贡献因子的差异性对制定最优策略提出了更高要求。另一方面, 考虑到实体属性的多维度特征, 例如在空气质量监测应用中, 污染、车流量、噪声等具有时空属性, 需要设计适用于多维属性的资源配置和实体选取模型以提高资源利用率和数据收集质量。

基于上述描述, 基于实体多属性的在线任务分配机制研究主要囊括以下三个方面:

- 1) 分析实体的空间属性对于资源配置影响, 研究如何设计任务分配算法, 在合理配置资源的前提下, 提高任务完成的质量。
- 2) 研究如何利用多维异构实体之间的相关性, 设计基于网络嵌入的多任务在线分配策略, 在有限预算条件下, 提升数据质量的准确性和完整性。
- 3) 研究演进预测结果和数据相关性对实体优化选取的影响, 设计任务在线调度策略, 以提升时空数据获取的高效性和可靠性。

2.2 研究目标

本项目将面向群智数据融合网络进行前瞻性研究, 从群智异构多维网络嵌入、用户数据



真值推断、基于实体多属性的在线任务分配的角度进行建模分析，系统探究关于群智数据融合网络的三个重要问题，每个研究内容具体目标如下：

群智场景的网络嵌入表征

- 1) 基于群智网络的多维度数据和群智场景的经验知识，构建不同语义的元路径集合，并实现快速有效的网络嵌入。
- 2) 基于演进群智网络的表征向量序列，结合网络平滑增长方式，实现对时变实体关系的准确预测。
- 3) 针对群智网络的带权特性和多样化的元路径集合，引入在随机游走过程中的交叉选择机制，得到承载多维度信息的嵌入向量。

基于时空相关的真值推断

- 1) 构建时间、空间相关性表达，通过耦合数据在时间和空间上的连续特征，将物理空间的时空属性映射到数据空间的数据相关性。
- 2) 基于感知环境建模用户能力，针对异质任务和多样化数据表征形式，分析数据质量与用户能力的关系，进而揭示数据噪声的产生机理，为构建推断算法提供理论依据。
- 3) 设计概率图模型推断任务真值，综合考虑物理空间时空相关性以及实体层面用户相关性，从概率最大化的角度准确推断真值。

基于实体多属性的在线任务分配机制研究

- 1) 将具有空间属性的在线任务分配代数问题图形化，设计基于图结构的确定性最优分配算法和激励机制，以提高资源配置的有效性。
- 2) 分析实体相关性和真实值推断对数据质量的影响，建立数据质量误差最小化模型，设计基于网络嵌入的多任务优先级分配策略，从而在有限预算条件下，提升数据质量准确度。
- 3) 基于演进预测信息和数据的时空相关性分析，结合数据收集的不同需求，设计在线实体优化选取策略，以提高群智计算的高效性。

2.3 拟解决的关键科学问题

通过上述三个方面的研究，本项目拟解决的问题如下：

- 1) 元路径优化设计问题。元路径能够作为网络中语义信息的有效载体，对网络表征结果具有关键影响。鉴于元路径的设计依赖于先验知识，通过人机协作能够有效降低这种依赖



关系以及元路径设计的不确定性。设计元路径组合策略，通过元路径交叉选择机制，提升表征向量对网络多维度信息的承载容量。

- 2) 数据相关性挖掘和真值推断问题。数据的时间、空间相关性能进一步提高真值推断的准确度，通过分析数据采集过程在时间、空间的连续特征可描述数据相关性信息，融合网络嵌入获取的用户相关性，从概率角度描述相关性对真值推断的影响，从多维度保证结果的准确性。
- 3) 基于实体多属性的在线任务分配机制设计问题。面向异构多态、复杂关联的群智环境，如何挖掘不同实体间的相关性特征，设计多任务优先级分配策略，实现计算效率与数据质量之间的折中，通过结合实体多维空间特征，促进资源的合理配置，提升群智计算的高效性和数据质量的准确性。

3. 拟采取的研究方案及可行性分析（包括研究方法、技术路线、实验手段、关键技术等说明）；

3.1 总体研究方案

本课题的总体研究方案框架如图 2 所示，研究群智数据融合网络中关键问题。针对群智系统物理空间中实体高维，异构，动态演进等特征，研究元路径构建策略，实现异构网络多维实体嵌入；进一步针对时变动态群智网络，探究网络演进规律，基于元路径集合，计算实体相似度，实现高维异构群智网络高效嵌入。为了提升群智系统数据质量，获取用户数据真实值，挖掘用户数据时空相关性，探究异质用户能力建模；进一步利用群智网络嵌入表征，融合数据与实体相关性，设计稳健的真值推断算法，提高推断准确度。为了实现群智系统高效精准任务分配需求，考虑任务空间相关性，构建基于图形化的空间任务最优分配策略；考虑时间维度，融合实体数据时空关联特征，构建基于实体相关性的在线任务分配策略；基于演进群智场景预测结果，考虑实体多维异构属性，构建基于多属性演进预测的在线任务分配策略。在理论上，实现基于网络嵌入表征的群智数据融合网络系统，综合提升群智系统数据计算及任务管理性能。

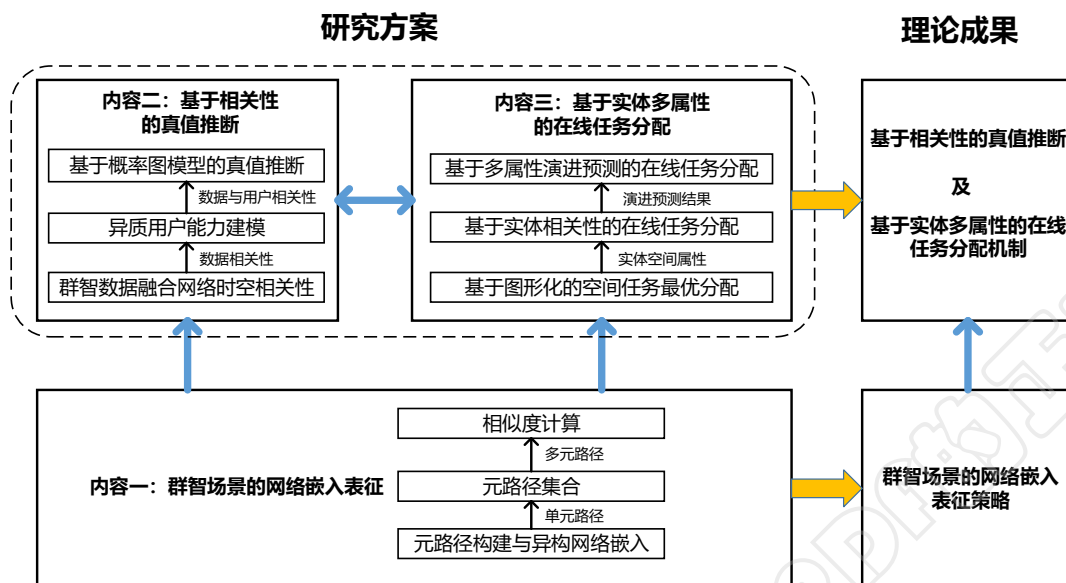


图 2. 总体研究方案框架

3.2 研究方案

3.2.1 群智场景的网络嵌入表征

随着通信方式的多样化,数据类型与社交联系日益复杂,导致群智网络中的用户特征淹没在庞大的冗余信息中,虽然网络嵌入能够提取网络结构中最有价值的信息,但是如何在群智场景中充分应用网络嵌入技术,是本课题的一个关键问题。考虑到元路径对网络语义信息的表征能力,我们拟使用基于元路径的嵌入算法挖掘群智网络中的用户相似度,并通过人工标定与神经网络相结合,筛选最符合群智场景的元路径集合,实现复杂群智的网络嵌入表征。

1) 元路径构建与异构网络嵌入

随着文本嵌入的发展,人们对网络嵌入的研究也有了长足的进步,通过随机游走的方式,能够将网络中的路径和节点,分别与文本数据中的上下文和单词对应起来,从而可以使用一般的文本嵌入的算法思想。但与传统的网络嵌入方式不同,基于元路径的异构网络嵌入,能够通过元路径将异构网络中特有的多种连接关系反应出来,即考虑到了异构网络结构中包含的场景语义信息,从而能够大幅度提高网络嵌入的准确性和可靠性,所以本项目拟采取基于元路径的异构网络嵌入来挖掘异构网络中的信息。

异构网络中,可以根据先验知识来设计合适的元路径。以购买网络为例,用 P、S、C 分别表示人、商店、和产品,就可以得到 P-C-S-C-P (在同一商店购买了同一商品), P-S-P (在



同一商店有过购买记录)等多条元路径,每条元路径都代表不同的信息。将这种元路径的模式设成一种统一的格式,有

$$\mathcal{P}: V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} \cdots V_{t-1} \xrightarrow{R_{t-1}} V_t,$$

其中 V_i 表示特定的节点种类。于是,就能够得到,随机游走在特定元路径模式下,下一跳的概率:

$$p(v^{i+1}|v_t^i, \mathcal{P}) = \begin{cases} \frac{1}{|N_{t+1}(v_t^i)|} & (v^{i+1}, v_t^i) \in E, \phi(v^{i+1}) = t+1 \\ 0 & (v^{i+1}, v_t^i) \in E, \phi(v^{i+1}) \neq t+1 \\ 0 & (v^{i+1}, v_t^i) \notin E \end{cases},$$

其中, ϕ 表示节点的种类, E 表示所有的连接, N_{t+1} 表示某个节点所有的 $t+1$ 类的邻居节点。当遍历网络中所有的节点,得到 N 条随机路径后,我们再利用 softmax 模型进行计算

$$p(c_t|v; \theta) = \frac{e^{x_{c_t} \cdot x_v}}{\sum_{u_t \in V_t} e^{x_{u_t} \cdot x_v}},$$

就可以得到每个节点嵌入后的向量。

2) 元路径集合

为了考虑真实世界中的不同关系,以及它们所表示的具体含义,设计多元化的元路径集合,来将经验知识转化为数据表示是必要且有效的。但是不同的元路径所承载的消息是不同的,甚至是对立的,无法任意的将这些元路径组合在一起,必须引入一种机制,使得算法能够有选择的对不同的元路径进行组合。

这里我们还是以商品购买的场景为例,令 P , S , C 分别表示人,商店和商品,我们能够构建 $P-S-P$, $P-C-S-C-P$, $P-C-P$ 三种元路径,分别表示两个人在同一商店有过购买记录,两个人在同一商店买过特定商品和两个人买过同一商品三种语义信息。

因为使用的是随机游走网络,所以可以通过控制游走的方向,来将三条元路径所表示的信息整合到一起。例如当游走到代表 P 的实体时,根据元路径模式,下一跳可以选择 S 或 C 。在以往的算法中,具体选择 S 还是 C 往往是完全随机的,但是考虑到群智网络的带权特性,这里需要引入对权重信息的考量。例如,可以直接将权值作为下一跳选择的概率值,权重越大的越有可能被考虑到。

通过多次迭代,就能够得到基于元路径集下的一条路径,这条路径就包含了上述三种语义信息,再将这条路径嵌入到向量空间,就实现了对多条元路径的有效组合。通过人工设定评分机制为不同元路径集合打分,多次迭代后,选取得分最高的元路径组合作为最终的组合策略。

3) 相似度计算



结合已有的相似度计算算法,选择最适合当前场景的算法,快速查找每个实体在向量空间中的最近邻。例如,已知一个多维空间 R^d , R^d 中的每一个向量都代表一个实体,则 R^d 中实体的集合可以称为实体集。给定实体集 N ,和一个实体 e , e 的最近邻就是任何实体 $e' \in E$ 满足 $\text{None-nearer}(N, e, e')$ 。其中, None-nearer 定义如下:

$$\text{None-nearer}(N, e, e') \Leftrightarrow \forall e'' \in E |e - e'| \leq |e - e''|.$$

上述的距离公式可以根据实际情况更改为其它的计算方式,例如可以使用余弦值、欧式距离或者点积等相似度计算方法。在得到每个实体的最近邻后,再结合群智网络的属性信息进行优化和筛选,得到最优的目标选择。

但是需要考虑的问题是,无论是哪种方法,都无法避免需要与所有的节点进行相似度计算,这对于大规模网络而言是十分耗时的,所以学习如何快速的在向量空间中寻找最近邻,是提升算法速度的有效途径。

网络嵌入后得到的向量因为保留了原本网络的结构信息,原本联系紧密的实体仍然会在向量空间中形成簇状聚类,所以不需要通过穷举的方式来进行逐点查询,可以根据这些聚类进行空间划分,缩小检索的范围。

例如,可以考虑利用k-d树的思想,对每一个向量建立索引,通过构造树形结构来索引每个节点的最近邻。k-d树本质上还是一种二叉树,通过均匀划分原则逐级递归展开,在这个过程中,所有的内节点都位于分隔平面上,而叶子节点都位于分隔区域内。构建好k-d树后,每次计算就可以通过向下搜寻和向上回溯两个阶段快速得到实体的最近邻。这种方法得到的最近邻可能不是空间上最紧的,但是能够大幅度提高算法的搜寻速度,从而减少算法的运行时间。

3.2.2 基于相关性的真值推断

群智数据融合网络的根本目的之一是利用收集数据,推断真值,为决策制定提供有效信息。真值推断依赖设计推断算法处理带噪声的数据从而获取真值结果。一方面,不同任务真值因物理空间时空属性影响在数据空间呈现相关性特征;另一方面,用户收集数据因实体层面相近而存在相关性;因此,融合数据相关性的推断算法能提高推断的准确性。为获取数据的时空相关性,通过分析数据的收集过程,映射时空临近关系到数据相关关系。考虑数据收集的噪声生成机理,建立适应感知任务的用户能力模型,刻画数据质量与用户能力的联系。利用最大似然概率,结合相关性特征,基于收集数据推断准确真值。

1) 群智数据融合网络时空相关性



与基于传感网和物联网的感知方式不同,群智感知以大量普通用户作为感知源,强调利用大众的广泛分布性、灵活移动性和机会连接性进行感知。在人类生产生活联系日益紧密的背景下,人类的行为在时空中存在越来越多的交织,因此基于人类行为的群智大数据也日益呈现越来越强的时空关联性。具体来说,群智用户同一时间在相邻地点和同一地点在相邻时间采集的感测数据分别具有空间和时间上的关联性。群智数据的时空关联性的应用极其广泛,如群智数据压缩优选、群智数据真实性发现等。非联合独立集被证明为一种强有效的数据关系挖掘手段从而被众多学者广泛采用,本项目拟采取非联合独立集分析的方案以实现群智数据时空关联性的高效挖掘。

群智数据在时间维度上具有相关性,如相邻时间的温度、空气质量测量等。时间关联性通常存在于一个给定的时间窗内,即感测数据只有在一个较小的时间段内具有相似性,所以可以通过构建非关联独立集来分析。假设 \mathcal{T}_i 是第 i 个时隙用户的集合, \mathcal{J}_i 是第 i 个独立集。总共有 $(p+1)(Q+1)$ 个时隙,用户数据的时间关联性在每个 $(p+1)$ 个连续时隙内都存在,则 $(p+1)$ 个非关联独立集可以表示为:

$$\{\mathcal{J}_i = \bigcup_{q=0}^Q \mathcal{T}_{i+(p+1)q}\}_{i=1}^{p+1}.$$

当 $p=1$ 时,当前时隙测量的数据仅仅与上一时隙测量的数据相关。在这种情况下,所有的时隙形成了两个非关联独立集,奇时隙集合 \mathcal{J}_1 和偶时隙集合 \mathcal{J}_2 ,从而可以进一步计算数据 i 和数据 j 之间的时间相关性 $T(i,j)$ 。

群智数据在空间维度上也具有相关性,如相邻加油站的油价、相邻地点的天气状况等。空间关联性具有局部性,下面通过举例来分析非关联独立集。假设所有区域内的所有用户都存在于一个网格地图中,并且被分为四个非关联独立集。令 $e_{(i,j)}$ 是网格地图中的第 i 行 j 列的位置, $p=\{1,2,\dots,P\}$ 和 $q=\{1,2,\dots,Q\}$ 是横纵坐标集合,那么四个独立集可以构建如下:

$$\mathcal{J}_1 = \{e_{(2p,2q-1)}\}_{p=1,q=1}^{P,Q}, \mathcal{J}_2 = \{e_{(2p-1,2q-1)}\}_{p=1,q=1}^{P,Q},$$

$$\mathcal{J}_3 = \{e_{(2p-1,2q)}\}_{p=1,q=1}^{P,Q}, \mathcal{J}_4 = \{e_{(2p,2q)}\}_{p=1,q=1}^{P,Q},$$

即偶数行奇数列形成独立集 \mathcal{J}_1 ,奇数行奇数列形成独立集 \mathcal{J}_2 ,奇数行偶数列形成独立集 \mathcal{J}_3 ,偶数行偶数列形成独立集 \mathcal{J}_4 ,类似时间相关性可以得到空间相关性 $L(i,j)$ 。

2) 异质用户能力建模

群智感知依赖用户完成任务,收集上传数据。而用户之间存在的能力差异使得其贡献的数据可靠性并不相同,甚至数据间彼此矛盾。因此,精准刻画用户质量是真值推断的基础。对用户异质性最直观的建模方法是利用概率参数表示用户数据的正确概率,该概率参数在



[0,1]区间, 值越大表示用户可靠性越高。本项目拟对不同难度等级的任务进行区分, 分别对连续和离散被测数据建模。

对于每个用户 i , K 个任务难度等级分别对应 K 个偏差-方差参数对 $(h_{i,k}, \sigma_{ik}^2)$ 。根基历史信息, 可以获取 $(h_{i,k}, \sigma_{ik}^2)$ 的先验分布信息, 具体来说 $h_{i,k}$ 服从正太分布:

$$p(h_{ik}|\mu_{ik}, \lambda_{ik}^2) = \frac{1}{\sqrt{2\pi}\lambda_{ik}} \exp[-\frac{(h_{ik}-\mu_{ik})^2}{2\lambda_{ik}^2}],$$

$\lambda_{i,k}$ 服从反向伽玛分布:

$$p(\sigma_{ik}^2|a_{i,k}, b_{i,k}) = \frac{b_{ik}^{a_{ik}}}{\Gamma(a_{ik})} (\sigma_{ik}^2)^{-a_{ik}-1} \exp\left[-\frac{b_{ik}}{\sigma_{ik}^2}\right],$$

其中 $a_{i,k}$, $b_{i,k}$ 是 σ_{ik}^2 的先验参数。

被测数据连续时, 用户收集数据 x_{ij} 与真值 z_j 和误差 e_{ij} 之间的关系为:

$$x_{ij} = z_j + e_{ij}.$$

单个用户收集数据的误差 e_{ij} 服从高斯分布, 则真值为 z_j 时其上传的数据为 x_{ij} 的条件概率服从:

$$p(x_{ij}|z_j, h_{ik}, \sigma_{ik}^2) = N(x_{ij}|z_j + h_{ik}, \sigma_{ik}^2) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp[-\frac{(x_{ij}-z_j-h_{ik})^2}{2\sigma_i^2}].$$

对于分类数据, 本项目拟建立基于贝叶斯方法的真值推断模型。任务的真值取自 l 个答案集合, 混淆矩阵的第 i 列表示当真值为第 i 个答案时, 用户数据在答案集合上的概率分布。通过对用户正确概率的动态修正, 即可有效去除数据中的噪声, 得出任务数据的真值。以二值数据为例, 混淆矩阵具体由四部分组成: 精密度、准确度、灵敏度以及特异度。精密度表示数据为正时的准确概率, 准确度为(所有)数据准确性, 灵敏度指示数据正真值也为正的概率, 特异度代表数据为负真值也为负的概率。

3) 基于概率图模型的真值推断

真值推断可以利用概率图模型描述用户数据的产生过程, 结合时间、空间相关性和用户社交属性, 通过最大化后验概率, 基于贝叶斯准则推断真值。

群智数据融合网络中存在 M 个用户 N 个任务, 用户 i 收集得到任务 j 的观测值为 x_{ij} , 任务真值为 z_j 。考虑到用户能力的异质性, 不同用户的观测值相对于真值的偏差不同, 本项目以偏差和方差的形式为代表说明用户模型对真值推断的影响。用户 i 的能力表示为 $e_{ik} = (h_{ik}, \sigma_{ik}^2)$, 其中 h_i 表示用户观测值的偏差, σ_i^2 表示用户观测值在偏差上下波动的方差。真值满足高斯分布:

$$p(z_j|\mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp[-\frac{(n_{ij}-\mu_j)^2}{2\sigma_j^2}].$$



根据异质用户能力建模可利用概率图模型能解释用户观测值与真值之间的关系，从概率层面阐述各项数据的产生过程，如图 3 所示：

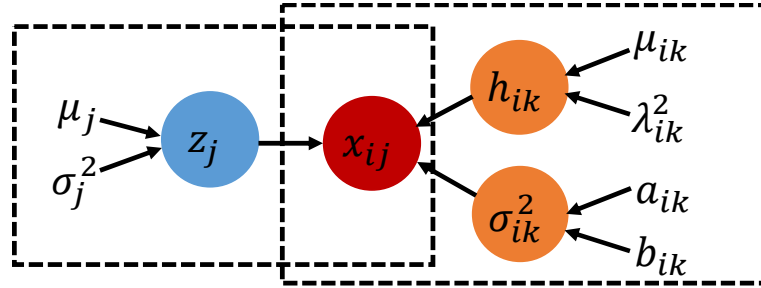


图 3：概率图模型

为了推断数据真值，依据概率图模型，从最大似然概率的角度可得用户观测值与任务真值的联合概率分布：

$$p(\mathbf{Z}, \mathbf{X}, \boldsymbol{\Theta}) = p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\Theta})p(\mathbf{Z})p(\boldsymbol{\Theta})$$

$$= \prod_j [p(z_j|\mu_j, \sigma_j^2)] \prod_i p(x_{ij}|z_j + h_{ik}, \sigma_{ik}^2) \prod_i [p(h_{ik}|\mu_{ik}, \lambda_{ik}^2) \times p(\sigma_{ik}^2|a_{ik}, b_{ik})]$$

另一方面，时间、空间相关性以及用户的社交数据直接影响真值的推断过程，从后验概率分布的角度来说，任务真值之间存在有时间、空间相关性，在时间和空间维度上距离更近的任务，其真值具有连续性，这一点已经在第 1) 部分提到，用户之间具有社交属性，从而用户观测值相依依赖，因此基于时间、空间相关性以及社交关系的推断目标为：

$$p'(\mathbf{Z}, \mathbf{X}, \boldsymbol{\Theta}) = p'(\mathbf{X}|\mathbf{Z}, \boldsymbol{\Theta})p'(\mathbf{Z})p(\boldsymbol{\Theta})$$

$$= \prod_j [p(z_j|\mu_j, \sigma_j^2) \times \alpha \prod_{j'} T(j, j') L(j, j')] \times \prod_i [p(x_{ij}|z_j + h_{ik}, \sigma_{ik}^2) \times \prod_{i'} S(i, i')] \prod_i [p(h_{ik}|\mu_{ik}, \lambda_{ik}^2) \times p(\sigma_{ik}^2|a_{ik}, b_{ik})]$$

其中 $T(j, j')$ 和 $L(j, j')$ 分别表示任务之间的时间、空间相关， $S(i, i')$ 表示用户之间的社交相关。

求解基于时空相关和社交关系的真值推断可利用最大似然概率以及 EM 算法，在每一步分别迭代真值和参数，直至收敛至稳态点。

3.2.3 基于实体多属性的在线任务分配机制研究

由于便携式设备的普及和无处不在的网络连接，群智系统已成为一种新型应用范式，广泛应用于环境感知、场景重现等。如何进行任务最优分配是一个亟待解决的关键问题。考虑到激励成本与数据质量之间存在折中，较高的成本可以允许更多用户参与任务分配，极大提升了数据质量。但对于特定应用而言，高成本往往是无法承受的。其次，不同的群智应用往往具有不同维度的实体属性，多实体属性往往具有一定的相关性，为实现最优任务分配带来



一定的困难。我们这里以群智感知应用为例，分别研究在有限预算下，单实体属性和多实体属性的感知任务在线分配问题。

1) 基于图形化的空间任务最优分配机制

某些感知应用仅存在单一实体属性，相比于多实体属性，由于不存在属性耦合，其最优任务分配将被最大程度地简化。这里考虑单一时间维度的感知属性，即感知数据仅是时间的函数。给定实体需求，传统的代数优化显然能够解决问题，但是考虑到感知数据的规模往往较大，从而不可避免地增加问题求解所需的时间，且随着数据规模的不断扩大，求解时间将以指数形式增长，为在线问题求解带来巨大困难。为此，需要设计合理的资源配置和实体匹配模型，从而尽可能在多项式时间内完成在线最优任务分配。

将优化问题从代数域转变至图形域是解决部分单属性最优任务分配的较优方法。具体地，对于时间维度的感知数据，我们可以将其抽象成图中的节点，将感知数据之间的联系抽象成连接不同节点的边。对于不同的限制条件，我们可以通过一定的等价代换，将限制因素（如有限预算等）映射为各条边的权重，从而将优化问题的可行域映射为一个有权图。进一步的，对于不同的优化目标，我们可以运用某些图论的经典算法来求解当前图对应的优化问题。

例如，在群智系统中利用出租车做群智道路车流量数据收集，可以将行将行程建模为途中的点，如果司机驾驶车辆能够满足两段乘客的行程之间时间的连接，那么将其建模为两个点之间边。考虑时间维度上的感知数据元组 $(t_i^p, t_i^d, l_i^p, l_i^d)$, $i = 1, 2, 3 \dots$ 。对于不同的元组 i 和 j ，如果满足 $t_i^d > t_j^p$ ，则称这两个元组是可连接的，即 $i \rightarrow j$ 。现在我们拥有这样的一批元组，对于最优任务分配而言，该问题等价于求解最少数量的时间链，使得所有元组都可被遍历。

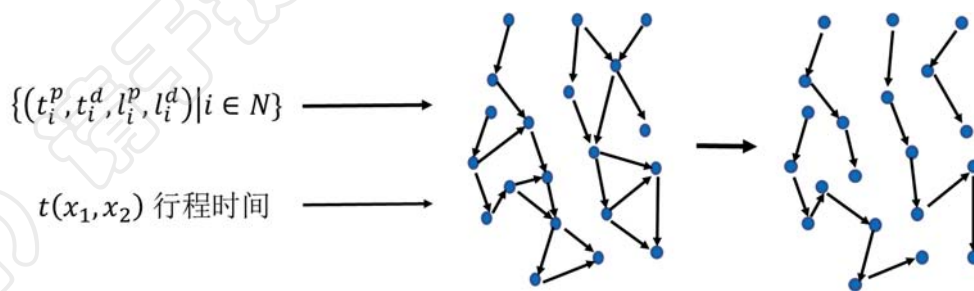


图 4：代数问题图形化示意图

该问题可以用传统的代数优化方法得到近似解，但效率较低。如果我们将元组抽象成节点，将元组之间的连接关系抽象成节点之间的边，那么将得到一个有向无环图。我们的目标是找到该有向无环图的最小路径覆盖数，即最少数量的时间链。由于有向无环图的最小路径覆盖问题是经典的数学问题，因此，将该优化问题从代数域转变为图形域求解将为后续的算法设计提供极大的便利。进一步的，如果我们对元组之间的连接性进行一定的限制，即不仅



仅考虑元组之间的时间先后，还考虑某些人为约束（如元组对应的实体之间是否可以共享等），则我们将根据这些限制条件重新构建各节点之间的边，并据此对边赋权值 w ，从而获得一个带权有向无环图。同样，对于带权有向无环图的最小路径覆盖问题，我们依然找到有性能保障的近似算法进行有效求解。

2) 基于实体相关性的在线任务分配机制

多维异构实体之间存在显著的相关性，考虑到已有任务分配方案大多假设不同实体间彼此相互独立，往往导致次优解的出现。这里研究基于实体相关性的在线任务分配机制。

将整个群智感知目标区域划分为 X 个同等大小不相交的感知单元，其中第 x 个感知单元用 x 表示。假设同一个感知单元内感知值一致。这里关注一个时间段内的数据质量，将其划分为 Y 个同等大小的时间槽，其中第 y 个时刻用 y 表示。当前许多感知应用中，感测变量往往具有多个实体感知属性，假定存在 A 个实体属性，其中 a 第 a 个属性用 a 表示。以城市空气质量检测系统为例，感知单元是感知值同一分布的一个邻域，感知时刻则为感知值更新的频率（如 1 小时），为了估计空气质量指数，用户需要收集一组与感测变量相关的实体属性（如 PM2.5, SO2, PM10）。定义矩阵 RS 表征所有的实际感知值，其中 RS^a 是属性 a 的实际感知值 $RS_{x,y}^a$ ，为实体属性 a 在感知单元 x 、时刻 y 的实际感知值。同样，我们定义所收集的感知值 CS ， CS^a ， $CS_{x,y}^a$ ，以及推断的感知值 IS ， IS^a ， $IS_{x,y}^a$ 。在此基础上，我们定义感知误差为推断的感知值与实际感知值间绝对误差的平均，特别地，属性 a 在感知单元 x 、时刻 y 的推断误差为 $SE_{x,y}^a = |IS_{x,y}^a - RS_{x,y}^a|$ 。将预算 B 定义为在每个时刻被选做任务分配的感知单元最大数目，在实际感知系统中，由于感知预算和资源受限性， $B < X$ 。

多维异构实体下的在线任务分配问题，其目标在于挖掘不同实体间的相关性，在给定预算 B 下，选择一组感知单元集合，最小化感测变量的聚合感知误差

$$\min_{S_y} \sum_{a=1}^A \Gamma_a \left(\frac{1}{Y_c} \sum_{y=1}^{Y_c} \frac{1}{X} \sum_{x=1}^X SE_{x,y}^a \right)$$

$$s.t. \quad |S_y| \leq B$$

其中， Y_c 表示当前感知时刻， S_y 为在时刻 y 被选做任务分配的感知单元集合。考虑到不同感知属性的值往往具有不同的范围，我们引入归一化方程 Γ_a 来归一化关于属性 a 的感知误差。进一步，定义任务优先级为在任务分配过程中感知单元被选中的顺序，并用优先级得分来量化，分数越高意味着优先级越高。通常，对于降低感知误差最有帮助的感知单元，在任务分配过程中应该具有最高优先级。

为了实时计算给定属性的每个感知单元的任务优先级，我们需要知道哪个单元的感知



值(如果被收集)对于减小感知误差是最有用的,在真实感知值事先不知道的情况下,该优化问题已被证明是一个 NP 完全问题。由于多维异构实体之间往往具有显著的相关性,为避免次优解的出现以及均衡计算量与效率,我们设计异构网络表征学习算法保证理论上可接受的近似率,基于元路径的随机行走,构建不同属性节点的异构邻域,挖掘低维潜在的多属性节点嵌入,得到考虑不同实体相关性的任务优先级。在此基础上,确定最优感知单元子集,实现聚合感知误差的最小化。

3) 基于多属性演进预测的在线任务分配机制

群智任务往往具有显著的时空属性,比如在空气质量监测应用中,污染、车流量、噪声等。此外,在线最优任务分配需要完备所有时隙的信息,需要对实体的未来位置等信息进行合理的预测和分析。因此,通过演进网络下的信息预测和时空相关性的分析,用户和任务可以在时空二维有效匹配以提高群智计算的高效性。故可将时空属性下的在线任务选取建模为有约束的多目标优化问题,并根据预测信息和历史数据设计在线调度算法。

将感兴趣的社区(城市)划分M个不同的单元 c_i ,单元之间的距离用矩阵S来表示,感兴趣的数据类型有K种,即有K种不同的任务。用 d_k 来表示任务的类型,且不同的数据类型具有不同的比重 p_k , $\sum_k p_k = 1$ 。不同类别的任务具有不同的时间影响函数 $h_k^T(t) \in [0,1], t \geq 0$,空间影响函数 $h_k^S(s) \in [0,1], s \geq 0$ 。假设区域内有N个实体,用 n_j 来表示,用矩阵 $B_{j,k}$ 来表示实体 n_j 是否能够完成任务k,矩阵 $G_{j,i}(t)$ 用来表示实体 n_j 在t时刻是否在单元 c_i 内,假设矩阵 $G_{j,i}(t)$ 是已知的。矩阵 $W_{j,k}(t)$ 表示在t时刻实体 n_j 是否能够收集k型数据,显然 $W_{j,k}(t) \leq B_{j,k}$ 。在演进网络信息整合预测下,可以获取矩阵 $G_{j,i}(t)$ 和 $B_{j,k}$ 的近似值。感知的目标是通过确定 $W_{j,k}(t)$,即实体在什么时刻什么地点收集何种数据以最优优化总体预期性能:

$$\max_{W[n]} E[\Gamma(X, U, Y)] = \gamma_1 E[X] + \gamma_2 E[U] - \gamma_3 E[Y],$$

$$\text{s.t. } W_{j,k}[n] \leq B_{j,k}, \forall j = 1, \dots, N, \forall k = 1, \dots, K,$$

$$E[D] \leq D_{quota}$$

其中, $E[X]$ 表示数据总体有效覆盖率期望; $E[U]$ 表示数据总体效用期望; $E[Y]$ 表示总体开销期望。我们用 $\omega_{i,k}[n] = [\omega_{i,1,k}[n], \dots, \omega_{i,N,k}[n]]$, $\omega_{i,j,k}[n] = W_{j,k}(n) \cdot G_{j,i}(n)$ 来表示 d_k 在 c_i 单元激活状态, 向量 $\mathbf{X}^0[n]$ 表示单个单元是否被至少一个实体的感知数据覆盖, 其元素计算为:

$$x_{i,k}^0[n] = x(\omega_{i,k}[n]) = 1 - \prod_{j=1}^N (1 - \omega_{i,j,k}[n]).$$

则总体有效时空覆盖率期望为:



$$x_{i,k}[n] = 1 - \prod_{v=1}^n \prod_{i=1}^M (1 - h_k^T(n-v) \cdot h_k^S(S_{i,i}) \cdot x_{i,k}^0[v]).$$

考虑所有数据类型的时空有效覆盖: $x[n] = \frac{1}{M} \cdot \sum_{k=1}^K \sum_{i=1}^M p_k \cdot x_{i,k}[n]$, $x[n':n] = \frac{1}{n-n'+1} \cdot \sum_{v=n'}^n x[v]$ 。与有效覆盖率类似, 我们可以得到总体效用的 U 。实体 n_j 的状态以及总的有效实体数量为:

$$y_j[n] = 1 - \prod_{k=1}^K (1 - W_{j,k}[n]), y[n] = \sum_{j=1}^N y_j[n].$$

目标函数中最后一个约束是由数据确定的数据配额有限的服务器资源和通信基础设施以保证所有数据都可以及时传输和处理。上述问题利用了演进网络预测信息和数据时空相关性分析, 且可以被证明是一个 NP 难的问题。在算法设计方面, 考虑实体收集数据的能力不同, 通过实体的时空历史数据收集和预测的位置信息将其能力进行量化评分, 设计高分先搜索的贪婪算法可以得到较好的在线实体及任务优选和匹配机制。在数据预算处理方面可以利用李雅普诺夫控制理论, 动态处理数据预算并保证其平均使用的有界性。

3.3 研究方案可行性分析

1) 具有一定的理论基础和前期积累

本课题所使用的理论工具主要是深度学习、贝叶斯网络、优化、统计等, 课题负责人及团队成员运用这些理论工具解决网络问题具有丰富的经验。曾运用相关理论工具研究过 Cellular-WiFi 双层网络中基于用户对 WiFi 网络的偏爱设计数据分流的契约理论机制、DTN 网络中基于用户时延容忍性设计数据分流机制等。课题团队在群智系统真值推断和激励机制设计方面做过一定理论工作, 包括群智答案推断分析, 群智激励机制、基于 BCH 码的群智系统多类标记方案等。此外在用户社交相关性方面也具有一定的研究基础, 基于用户之间的实时交互信息, 结合数学统计信息设计用户间社交关系强弱的获取机制。在这三个方面的研究均取得了一定的成果, 发表在多个国际知名会议, 包括 IEEE INFOCOM, IEEE GLOBECOM, IEEE ICC 等。

2) 研究方案可行, 相关理论与技术成熟

目前, 基于网络嵌入的数据分析研究不断深入, 为我们提供了强大的理论支持。本项目研究方案制定遵循研究内容先后顺序进行研究, 采用网络嵌入对群智场景进行表征, 不仅能提取群智网络中实体之间相关性, 而且降低了高维数据处理复杂度, 提升了群智系统的性能; 依赖于群智系统实体之间相似度表征, 本课题探索了基于数据时空性和实体相似度的真值计算机制以及基于实体多属性的在线任务分配机制, 从数据处理计算和实体任务分配两



个层次实现了群智数据融合网络性能优化。三个方面相辅相成、环环相扣，采用了回溯完善的研究模式。新思路建立在已有研究基础之上，结合相关领域内的新进展和新方法而提出，研究目标的确立结合实际，研究内容的制定逻辑清晰，研究方案的设计科学合理。

3) 人员配置合理，具有扎实的相关工作基础和完善的技术支持

课题组中的中青年教师都具有深厚的科研学术沉淀和丰富的项目承担经验，其他主要成员也均是优秀的研究生和博士生，数学基础理论扎实，在基础科研工作中曾经取得过一些成果，发表在 IEEE JSAC, IEEE TWC, IEEE TMC, IEEE TVT, IEEE TPDS, IEEE INFOCOM, ACM SIGKDD, ACM SIGIR, AAAI, IJCAI, IEEE GLOBECOM, IEEE ICC, ACM MOBIHOC 等国际知名期刊和会议上。在项目实践方面也有很好的表现，具有参与过自然基金重点和面上项目的经验。综上所述，本课题组具有良好的学术积累、明确的研究方向、可行的技术路线，完全可以在计划时间内完成该项研究，取得预期的研究成果。

4. 本项目的特色与创新之处；

创新点：

- 1) **提出基于群智场景元路径的网络嵌入算法。**结合多维度时空数据和场景先验知识，设计元路径集合，通过交叉选择机制得到承载不同维度信息的实体向量，并能基于簇内的局部搜索，快速寻找最近邻，实现对目标的相关性比较。
- 2) **提出融合数据与实体相关性的贝叶斯推测算法：**映射物理空间临近关系到数据空间相关性，充分表现时空相关性的影响，综合考虑异质用户数据差异性，通过概率图模型描述清晰描述数据产生过程，融合时空相关和网络嵌入精确推断真值。
- 3) **提出基于实体多属性的在线任务分配机制。**挖掘实体属性的异构多态、复杂关联特征，揭示实体间潜在的低维嵌入，促进异构时空任务分配和资源合理配置，增强群智网络区域性管理能力。

特色：

本课题以网络嵌入、深度学习、优化、统计等理论为工具，对群智数据融合网络中的的关键问题进行前瞻性研究，具备一定的理论挑战性。与目前已有工作相比，采用网络嵌入对群智网络场景进行表征，不仅可以准确提取网络实体之间相关性，而且降低了高维特征处理复杂度，提升了群智系统的效率；基于群智场景网络嵌入表征的基础，融合数据时空相关性和用户能力异质性，实现群智网络真值精准推断；基于网络嵌入的实体相似度表征和数据真



值推断进一步提升了群智系统中任务管理能力，因此本项目探索了群智系统中基于实体多属性的在线任务分配机制，挖掘时间的异构多态属性和复杂关联特征，促进异构时空任务分配和资源合理配置。

5. 年度研究计划及预期研究结果（包括拟组织的重要学术交流活动、国际合作与交流计划等）。

5.1 年度研究计划

2020 年 1 月~ 2020 年 12 月	实现基于单元路径的网络群智场景嵌入表征,分析群智网络多维度时空数据,基于场景的先验知识,探索元路径设计方案;分析多时刻群智网络的嵌入表征,基于动态群智网络的表征向量序列,结合网络平滑演进框架,实现对动态实体关系的准确预测。
2021 年 1 月~ 2021 年 12 月	实现基于多元路径的网络群智场景嵌入表征。针对群智网络的带权特性和多样化的元路径集合,研究随机游走过程中的交叉选择策略,提升嵌入结果承载不同维度,得到承载多维度信息的嵌入向量。
2022 年 1 月~ 2022 年 12 月	设计融合数据与实体相关性的贝叶斯推断算法。研究数据相关性的时空表征,解释物理空间到数据相关性的映射关系,分析数据相关性的计算逻辑。探索用户能力的异质性模型,刻画用户模型对网络嵌入分析的影响。设计概率图模型推断任务真值,从概率最大化的角度准确推断真值。
2023 年 1 月~ 2023 年 12 月	设计基于实体多属性的在线任务分配机制。将具有空间属性的在线任务分配的代数问题图形化,设计基于图结构的确定性最优分配算法和激励机制;分析实体相关性和真实值推断对数据质量的影响,建立数据质量误差最小化模型,设计基于网络嵌入的多任务优先级分配策略;基于演进预测信息和数据的时空相关性分析,设计在线实体优化选取策略。

5.2 预期研究成果

- 完成群智场景的网络嵌入表征,设计元路径组合策略,通过元路径交叉选择机制,提升表征向量对网络多维度信息的承载容量。



- 完成群智系统数据相关性挖掘和真值推断，构建可描述数据时空相关性信息，融合网络嵌入获取的用户相关性，从概率角度描述相关性对真值推断的影响，从多维度保证结果的准确性。
- 构建基于实体多属性的在线任务分配机制，通过结合实体多维预测结果，权衡成本与数据质量之间的折中，促进资源的合理配置，提升群智计算的高效性和数据质量的准确性。
- 在国际知名刊物和国内核心期刊上发表论文 6~10 篇，在有影响力的国际会议上发表论文 6~10 篇。
- 申请国家发明专利 4~5 项。
- 培养研究生 6~8 名。
- 邀请国际知名研究机构相关领域学者访问 2-4 次。

（二）研究基础与工作条件

1. 研究基础（与本项目相关的研究工作积累和已取得的研究工作成绩）；

（1）群智数据融合网络研究基础

课题组在群智系统数据分析，激励机制，以及任务分配理论研究方面具有一定的基础。研究了成本受限情况下的群智系统多层次标识任务激励机制，使得系统在多项式时间复杂度下达到高性能，成果发表在 IEEE JSAC 2017 [1]。在具有社交相关性的群智系统多资源分配激励机制研究中，考虑到资源公平性，根据博弈论提出了资源分享算法，实现了资源分配公平性和效率的均衡，研究成果发表在 IEEE JSAC 2017 [2]。研究了群智网络的覆盖性能与检测范围之间的关系，成果被 IEEE TMC [3] 录用。研究了基于地理位置隐私保护相关性的群智系统激励机制，采用 k-匿名方法来保护系统用户地理位置隐私，并得到了用户隐私保护和系统性能的均衡，成果发表在 IEEE TWC 2017 [4]。考虑了交通路况实时信息获取真实性问题，我们首先为每个信息贡献者建立了声誉系统，以提升用户可靠性，同时进一步设立了激励机制来刺激可靠用户提供更加准确信息，成果发表在 IEEE TITS 2017 [5]。针对蜂



窝网络中时延负载均衡问题,我们首先根据契约理论将其建模为一个寡头市场,我们设计了激励机制以根据用户的时延和价格敏感选择服务,大大缓解了蜂窝网络中国的拥塞问题,成果发表在 IEEE TWC 2016 [6]。我们探究了具有社交相关性的无线网络容量,针对无线网络中多跳通信干扰问题,采用定向天线技术来提升网络容量,在密集社交社交无线网络中,网络容量可以达到最优,成果发表在 IEEE TCOM 2017 [7]。在具有地理相关性的群智系统任务分配中,考虑到地理位置的影响和用户资源的公平性,基于 Lyapunov 算法动态优化分配方案,有效提升了平台效用和公平性,成果发表在 IEEE INFOCOM 2017 [8]。提出了一种考虑预算限制,针对标记型群智计算平台的激励机制,以鼓励群智工人参与群智计算,成果发表在 IEEE INFOCOM 2015 [9]。基于群智推断和激励机制,提出了无线信号,如 WIFI 信号,指纹采集点的分布策略,研究群智环境中在不同情况下测量点的分布模式以及它对室内定位精度的影响,给出了定位精度的理论界限,成果发表于 IEEE INFOCOM 2015 [10]。

本项目前期工作发表的相关论文

- [1] Xiaoying Gan, X. Wang, W. Niu, G. Hang, X. Tian, X. Wang, J. Xu, “[Incentivize Multi-Class Crowd Labeling Under Budget Constraint](#)”, **IEEE Journal on Selected Areas in Communications**, vol. 35, no. 4, pp: 893-905, 2017. (CCFA 类)
- [2] Xiaoying Gan, Y. Li, W. Wang, L. Fu, X. Wang, “[Social Crowdsourcing to Friends: An Incentive Mechanism for Multi-Resource Sharing](#)”, **IEEE Journal on Selected Areas in Communications**, vol. 35, no. 3, pp: 795- 808, 2017. (CCFA 类)
- [3] Xiaoying Gan, Z. Zhang, Y. Fu, X. Wu, X. Wang, “[Unraveling Impact of Critical Sensing Range on Mobile Camera Sensor Networks](#),” **IEEE Transactions on Mobile Computing**, DOI: 10.1109/TMC.2019.2901478, 2019. (CCFA 类)
- [4] X. Wang, Z. Liu, X. Tian, Xiaoying Gan; Y. Guan, X. Wang, “[Incentivizing Crowdsensing With Location-Privacy Preserving](#)”, **IEEE Transactions on Wireless Communications**, vol. 16, no. 10, pp: 6940–6952, 2017. (CCF B 类)
- [5] X. Wang, R. Jia, X. Tian, Xiaoying Gan, L. Fu, X. Wang, “[Location-Aware Crowdsensing: Dynamic Task Assignment and Truth Inference](#),” **IEEE Transactions on Mobile Computing**, 2018. DOI: 10.1109/TMC.2018.2878821. (CCFA 类)
- [6] X. Wang, J. Zhang, X. Tian, Xiaoying Gan, Y. Guan, X. Wang, “[Crowdsensing-Based Consensus Incident Report for Road Traffic Acquisition](#)”, **IEEE Transactions on Intelligent**



- Transportation Systems**, vol. PP, no. 99, pp. 1 – 12, 2017. (CCF B 类)
- [7] Y. Li, J. Zhang, **Xiaoying Gan**, L. Fu, H. Yu, X. Wang, “[A Contract-Based Incentive Mechanism for Delayed Traffic Offloading in Cellular Networks](#)”, **IEEE Transactions on Wireless Communications**, vol. 15, no. 8, pp: 5314-5327, 2016. (CCF B 类)
- [8] Z. Qin, **Xiaoying Gan**, J. Wang, L. Fu, X. Wang, “[Capacity of Social-aware Wireless Networks with Directional Antennas](#)”, **IEEE Transactions on Communications**, vol. 65, no. 9, pp. 4831–4844, 2017. (CCF B 类)
- [9] X. Wang, R. Jia, X. Tian, **Xiaoying Gan**, “[Dynamic Task Assignment in Crowdsensing with Location Awareness and Location Diversity](#)”, **IEEE INFOCOM**, Honolulu, USA, 2018. (CCF A 类)
- [10] Q. Zhang, Y. Wen, X. Tian, **Xiaoying Gan**, X. Wang, “[Incentivize Crowd Labeling under Budget Constraint](#)”, **IEEE INFOCOM**, Hongkong, China, 2015. (CCF A 类)
- [11] K. Sheng, Z. Gu, X. Mao, X. Tian, W. Wu, **Xiaoying Gan**, X. Wang, “[The Collocation of Measurement Points in Large Open Indoor Environment](#)”, **IEEE INFOCOM**, Hongkong, China, 2015. (CCF A 类)

(2) 数据特征挖掘与分析研究基础

课题组成员近5年来积累了深度学习和数据挖掘方面的研究工作，在数据科学和人工智能的国际知名期刊和会议上发表代表性第一作者或通讯作者论文25篇(其中SCI 一区论文1篇，CCF A类论文12篇，CCF B类论文8篇，CCF C类论文2篇)，代表性合作论文9篇(CCF A类论文2篇，CCF B类论文3篇，CCF C类论文1篇)。

课题组成员在SIGIR 2017发表三个strong accept的满分论文“IRGAN: A Minimax Game for Unifying Generative and Discriminative Information Retrieval Models”的对抗训练模型工作、在AAAI 2017发表的题为“SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient”的生成式对抗网络训练工作以及在AAAI 2018发表的题为“Aggregating Crowd Wisdoms with Label-aware Autoencoders”的群智系统标签汇聚学习工作。这些前期的研究作为课题组在相关领域积攒了学术经验和工程基础，为申为申报的课题奠定了坚实的理论基础和技术保障。



(3) 申请人及主要参与者具有承担科研项目的经验

项目申请人及主要参与者承担群智系统与机器学习相关的项目包括：国家自然科学基金面上项目——面向社交网络的无线资源管理机制研究（61672342）；国家自然科学基金青年项目——基于多域离散数据的深度学习(61702327)；国家自然科学基金青年项目——认知接入中能量效率问题的研究（61102052）。以上项目为本项目的研究打下深厚的基础。

2. 工作条件（包括已具备的实验条件，尚缺少的实验条件和拟解决的途径，包括利用国家实验室、国家重点实验室和部门重点实验室等研究基地的计划与落实情况）；

项目申请人和成员所在的上海交通大学智能物联网研究中心是在于全院士领导下，陈贵海教授担任主任，王新兵教授担任常务副主任的科研团队。隶属于上海交通大学电子信息与电气工程学院计算机科学与信息工程博士后流动站下的一个重要的研究机构。该中心目前已经形成了一个物联网与数字媒体相关的学科基地和科研大平台，实验室总面积超过1500平方米，拥有先进的通信网络技术研究开发设备和开发软件，包括物联网软件和硬件开发平台、IP核及其开发系统、多格式高清晰度录像机、B3G/4G、信道衰落仿真仪、500MHz数字存储示波器、射频信号发生器、任意基带波形发生器、数字逻辑分析仪、实时码流分析仪、统计信号处理服务器、高清晰度监视器等重要设备，固定资产总值2200多万元。通过从海外引进高端人才，包括国家千人计划、青年千人计划等，充实教授队伍，现在拥有六名博士生导师，八名硕士生导师，十几名年轻教师骨干，有四名在站博士后，二十多名博士生，六十多名硕士生，并以无线通信网络的传输与控制技术见长，是学校211工程和985工程重点资助对象，先后承担过多个国家863、973项目和国家自然科学基金项目以及国际、国内合作开发项目。在国际顶级期刊发表学术论文120余篇（如IEEE JSAC, IEEE TIT, IEEE TON, IEEE TMC, IEEE TCOM等），会议论文250余篇，引用次数超过8000次，获取多项国内外学术荣誉及奖项，包括中国国家专利金奖，国家科技进步二等奖、IEEE通信学会亚太杰出青年研究者奖等。同时和多所国内知名无线通信研究高校保持合作关系，并与国际知名的无线网络专家保持长期的合作与交流关系。

3. 正在承担的与本项目相关的科研项目情况（申请人和项目组主要参与者正在承担的与本项目相关的科研项目情况，包括国家自然科学基金的项目和国家其他科技计划项目，要注明项目的名称和编号、



经费来源、起止年月、与本项目的关系及负责的内容等);

国家自然科学基金面上项目: 面向社交网络的无线资源管理机制研究 (61672342)

经费来源: 国家财政拨款

项目起止年月: 2017.01-2020.12

课题负责人: 甘小莺

与本项目的关系: 基于群智系统真实值推测及任务分配激励机制设计是本项目的重要基础之一。

国家自然科学基金青年项目: 基于多域离散数据的深度学习 (61702327)

经费来源: 国家财政拨款

项目起止年月: 2018.01-2020.12

课题负责人: 张伟楠

与本项目的关系: 标签数据的深度学习是本项目的重要基础之一。

4. 完成国家自然科学基金项目情况 (对申请人负责的前一个已结题科学基金项目 (项目名称及批准号) 完成情况、后续研究进展及与本申请项目的关系加以详细说明。另附该已结题项目研究工作总结摘要 (限 500 字) 和相关成果的详细目录)。

国家自然科学基金青年基金项目: 认知接入中能量效率问题的研究 (61102052)

项目起止年月: 2012.01-2014.12

课题负责人: 甘小莺

完成情况: 达到项目预期要求, 并超额完成任务。

工作总结摘要: 认知技术可以改善频谱利用率, 缓解频谱资源稀缺的矛盾。本课题对认知接入中能量效率问题进行研究, 以认知接入的能耗模型为基础, 分别从用户和系统两个层面, 对认知接入、频谱共享和传输机制等问题进行分析和优化, 并深入结合经济学原理, 对接入用户的激励问题进行分析 and 研究。研究成果包括:

在认知用户随机接入多信道时, 提出了能效驱动的多用户接入最优门限, 使得用户能够根据对信道状态的观测判定是否接入该信道, 以达到系统能量效率最大化的目的, 在此基础上, 分析了非合作博弈情况下用户接入的能效损失比, 为分布式系统设计提供理论参考; 在



多个初级用户与多个次级用户共存,考虑用户自私性的协同频谱共享市场中,将认知中继网络建模为一个匹配市场模型,在不完全信息场景中,推导了生成稳定匹配的充分和必要条件,在此基础上,提出了基于收益增加法的分布式匹配算法,并证明了在部分不完全信息场景下,所提算法能够收敛到帕累托最优均衡点。

本课题共发表 31 篇学术论文,其中期刊论文 12 篇,会议论文 19 篇;属于 SCI 期刊源的国际刊物论文 9 篇, EI 源论文 30 篇。论文发表的国际知名刊物如 IEEE TPDS, TCOM, TWC, TVT 等,有影响力的重要国际学术会议如 IEEE INFOCOM, GLOBECOM, ICC, WCNC 等;申请发明专利 7 项,公开 7 项,授权 2 项。

代表性研究成果列表

- [1] Xiaoying Gan and B. Chen, “A Novel Sensing Scheme for Dynamic Multichannel Access”, IEEE Trans. Vehicular Technology, vol. 61, no. 1, pp. 208-221, Jan. 2012.
- [2] Xiaoying Gan, M. Xu, and H. Li, “Energy Efficient Sequential Sensing in Multi-User Cognitive Ad Hoc Networks: A Consideration of an ADC Device”, Journal of Communications and Networks, vol. 14, no. 2, pp. 188-194, Apr. 2012.
- [3] X. Feng, G. Sun, Xiaoying Gan, F. Yang, X. Tian, X. Wang, and G. Mohsen, “Cooperative Spectrum Sharing in Cognitive Radio Networks: A Distributed Matching Approach”, IEEE Trans. Communications, vol. 62, no. 8, pp. 2651-2664, Aug. 2014.
- [4] H. Li, Xiaoying Gan, S. Chen, and X. Feng, “Multi-channel Spectrum Sensing in Cognitive Ad-hoc Networks: an Energy-Efficient Manner”, Journal of Shanghai Jiaotong University (Science), vol. 18, no. 5, pp. 513-519, Oct. 2013.
- [5] H. Zheng, F. Yang, X. Tian, and Xiaoying Gan, “Data Gathering with Compressive Sensing in Wireless Sensor Networks: A Random Walk Based Approach”, IEEE Trans. Parallel and Distributed Systems, vol. 26, no. 1, pp. 35-44, Jan. 2015.
- [6] Y. Qin, J. Zheng, X. Wang, H. Luo, H. Yu, X. Tian, and Xiaoying Gan, “Opportunistic Scheduling and Channel Allocation in MC-MR Cognitive Radio Networks”, IEEE Trans. Vehicular Technology, vol. 63, no. 7, pp. 3351-3368, Sept. 2014.
- [7] G. Sun, X. Feng, X. Tian, Xiaoying Gan, Y. Xu, X. Wang, and G. Mohsen, “Coalitional Double Auction for Spatial Spectrum Allocation in Cognitive Radio Networks”, IEEE Trans. Wireless Communications, vol. 13, no. 6, pp. 3196-3206, Jun. 2014.



- [8] C. Li, Z. Liu, X. Geng, M. Dong, F. Yang, **Xiaoying Gan**, X. Tian, and X. Wang, “[Two Dimension Spectrum Allocation for Cognitive Radio Networks](#)”, IEEE Trans. Wireless Communications, vol. 13, no. 3, pp. 1410-1423, Mar. 2014.
- [9] X. Feng, H. Wang, and X. Wang, “[A Game Approach for Cooperative Spectrum Sharing in Cognitive Radio Networks](#)”, Wireless Communications and Mobile Computing, vol. 15, no. 3, pp. 538-551, Mar. 2013.
- [10] K. Sheng, Z. Gu, X. Mao, X. Tian, W. Wu, **Xiaoying Gan**, and X. Wang, “[The Collocation of Measurement Points in Large Open Indoor Environment](#)”, in Proc. IEEE INFOCOM, Apr. 2015.
- [11] Q. Zhang, Y. Wen, X. Tian, **Xiaoying Gan**, and X. Wang, “[Incentivize Crowd Labeling under Budget Constraint](#)”, in Proc. IEEE INFOCOM, Apr. 2015.
- [12] S. Chen, **Xiaoying Gan**, X. Feng, X. Tian, W. Wu, and J. Liu, “[Markov Approximation for Multi-RAT Selection](#)”, in Proc. IEEE ICC, Jun. 2015.
- [13] Y. Hu, X. Wang, and **Xiaoying Gan**, “[Critical Sensing Range for Mobile Heterogeneous Camera Sensor Networks](#)”, in Proc. IEEE INFOCOM, Apr. 2014.
- [14] X. Xiao, X. Tian, **Xiaoying Gan**, and X. Wang, “[Multi-Class Labeling with BCH codes for Mobile Crowdsensing](#)”, in Proc. IEEE GLOBECOM, Dec. 2014.
- [15] K. Sheng, Z. Gu, X. Mao, X. Tian, **Xiaoying Gan**, and X. Wang, “[Answer Inference for Crowdsourcing based Scoring](#)”, in Proc. IEEE GLOBECOM, Dec. 2014.
- [16] J. Huang, **Xiaoying Gan**, and X. Feng, “[Multi-Armed Bandit Based Opportunistic Channel Access: a Consideration of Switch Cost](#)”, in Proc. IEEE ICC, Jun. 2013.
- [17] 冯心欣, 王路洋, 王新兵, **甘小莺**, 田晓华, “[认知无线电中基于主用户估测的频谱合作共享方法](#)”, 2014/9/24, 31, 中华人民共和国国家知识产权局, CN201410282932.X, 专利
- [18] **甘小莺**, 龙鑫, 张炜, 周元, “[感知无线电中次要用户的频谱接入方法及装置](#)”, 2013/3/20, 44, 中华人民共和国国家知识产权局, CN201210468099.9, 专利

(三) 其他需要说明的问题

1. 申请人同年申请不同类型的国家自然科学基金项目情况 (列



明同年申请的其他项目的项目类型、项目名称信息，并说明与本项目之间的区别与联系。

无。

2. 具有高级专业技术职务（职称）的申请人或者主要参与者是否存在同年申请或者参与申请国家自然科学基金项目的单位不一致的情况；如存在上述情况，列明所涉及人员的姓名，申请或参与申请的其他项目的项目类型、项目名称、单位名称、上述人员在该项目中是申请人还是参与者，并说明单位不一致原因。

无。

3. 具有高级专业技术职务（职称）的申请人或者主要参与者是否存在与正在承担的国家自然科学基金项目的单位不一致的情况；如存在上述情况，列明所涉及人员的姓名，正在承担项目的批准号、项目类型、项目名称、单位名称、起止年月，并说明单位不一致原因。

无。

4. 其他。

无。



甘小莺 简历（申请人）

上海交通大学，电子信息与电气工程学院，副教授

申请人是上海交通大学“智能物联网研究中心”副主任，主持国家自然科学基金青年项目、面上项目，参与重点项目与杰青项目，已授权国家发明专利30余项（美国专利1项）。作为负责人，获得飞利浦公司、波音公司及华为公司科研项目资助。近五年来，申请人围绕群智系统原理与无线网络激励机制设计开展研究，针对群智系统大规模移动用户参与的激励机制设计、隐私保护与系统性能的均衡以及用户声誉系统构建等重要问题，取得了一些创新性的学术成果，相关研究成果在IEEE期刊发表长文20余篇。获得军队科技进步三等奖(2007)“高效Turbo乘积码编译码技术研究与实现”、上海市技术发明二等奖(2012)“基于认知的无线传感器网络在工业信息话中的关键技术与核心系统”和上海市科技进步三等奖(2018)“基于云WiFi的无线城市运营平台”。

代表性论著

1. Xiaoying Gan, X. Wang, W. Niu, G. Hang, X. Tian, X. Wang, J. Xu, “[Incentivize Multi-Class Crowd Labeling Under Budget Constraint](#),” **IEEE Journal on Selected Areas in Communications**, vol. 35, issue 4, pp: 893-905, 2017. (CCFA 类)
2. Xiaoying Gan, Y. Li, W. Wang, L. Fu, X. Wang, “[Social Crowdsourcing to Friends: An Incentive Mechanism for Multi-Resource Sharing](#),” **IEEE Journal on Selected Areas in Communications**, vol. 35, issue 3, pp: 795-808, 2017. (CCFA 类)
3. Xiaoying Gan, Z. Zhang, Y. Fu, X. Wang, “[Unraveling Impact of Critical Sensing Range on Mobile Camera Sensor Networks](#),” **IEEE Transactions on Mobile Computing**, DOI: 10.1109/TMC.2019.2901478, 2019. (CCFA 类)
4. Xiaoying Gan, C. Feng, Z. Qin, G. Zhang, H. Wu, L. Fu, X. Wang, H. Ma, “[Relay Assisted Multicast with Markov Mobility](#),” **IEEE Transactions on Network Science and Engineering**, Accepted, 2019. DOI:10.1109/TNSE.2019.2902801.
5. Xiaoying Gan, B. Chen, “[A Novel Sensing Scheme for Dynamic Multichannel Access](#),” **IEEE Transaction on Vehicular Technology**, vol. 61, no. 1, pp. 208-221, Jan. 2012.
6. Y. Li, W. Dai, J. Bai, Xiaoying Gan, J. Wang, X. Wang, “[An Intelligence-Driven Security-Aware Defense Mechanism for Advanced Persistent Threats](#),” **IEEE Transactions on Information Forensics & Security**, vol. 14, issue 3, pp:646-661, 2019. (CCFA 类)



7. X. Wang, R. Jia, X. Tian, **Xiaoying Gan**, L. Fu, X. Wang, “[Location-Aware Crowdsensing: Dynamic Task Assignment and Truth Inference](#),” **IEEE Transactions on Mobile Computing**, 2018. DOI: 10.1109/TMC.2018.2878821. (CCF A 类)
8. X. Wang, R. Jia, X. Tian, **Xiaoying Gan**, “[Dynamic Task Assignment in Crowdsensing with Location Awareness and Location Diversity](#),” in **Proc. Of IEEE INFOCOM**, Hi, USA, April. 2018. (CCF A 类)
9. Q. Zhang, Y. Wen, X. Tian, **Xiaoying Gan**, X. Wang, “[Incentivize Crowd Labeling under Budget Constraint](#),” in **Proc. Of IEEE INFOCOM**, Hong Kong, China, April. 2014. (CCF A 类)
10. Z. Qin, **Xiaoying Gan**, J. Wang, L. Fu, X. Wang, “[Capacity of Social-aware Wireless Networks with Directional Antennas](#),” **IEEE Transactions on Communications**, vol. 65, no. 11, pp: 4831-4844, Aug. 2017. (CCF B 类)



甘小莺 简历

上海交通大学，电子信息与电气工程学院，副教授

教育经历（从大学本科开始，按时间倒序排序；请列出攻读研究生学位阶段导师姓名）：

- (1) 2000.9 - 2005.11, 上海交通大学, 通信与信息系统, 博士, 导师: 宋文涛
- (2) 1996.9 - 2000.7, 上海交通大学, 通信工程, 学士, 导师:

科研与学术工作经历（按时间倒序排序；如为在站博士后研究人员或曾有博士后研究经历，请列出合作导师姓名）：

- (1) 2012.1-至今, 上海交通大学, 电子工程系, 副教授
- (2) 2009.4-2010.4, 美国加州大学圣迭戈分校, 电子与计算机工程, 访问学者
- (3) 2005.11-2011.12, 上海交通大学, 电子工程系, 讲师

曾使用其他证件信息（申请人应使用唯一身份证件申请项目，曾经使用其他身份证件作为申请人或主要参与者获得过项目资助的，应当在此列明）：

主持或参加科研项目（课题）情况（按时间倒序排序）：

- 1、国家自然科学基金面上项目，61672342，面向社交网络的无线资源管理机制研究，2017/01 - 2020/12，63 万元，在研，主持。
- 2、国家自然科学基金面上项目，61671478，卫星通信系统的信道建模理论与验证方法研究，2017/01 - 2020/12，60 万元，在研，参加。
- 3、国家自然科学基金重大研究计划培育项目，91438115，动态空间信息网络的传输容量与关键技术，2015/01 - 2017/12，80 万元，已结题，参加。
- 4、国家自然科学基金青年项目，61102052，认知接入中能量效率问题的研究，2012.1 - 2014.12，28 万元，已结题，主持。

代表性研究成果和学术奖励情况

（请注意：①投稿阶段的论文不要列出；②对期刊论文：应按照论文发表时作者顺序列出全部作者姓名、论文题目、期刊名称、发表年代、卷（期）及起止页码（摘要论文请加说明）；③对会议论文：应按照论文发表时作者顺序列出全部作者姓名、论文题目、会议名称（或会议论文集名称及起止页码）、会议地址、会议时间；④应在论文作者姓名后注明第一/通讯作者情况：所有共同第一作者均加注上标“#”字样，通讯作者及共同通讯作者均加注上标“*”字样，唯一第一作者且非通讯作者无需加注；⑤所有代表性研究成果和学术奖励中本人姓名加粗显示。）



按照以下顺序列出：①代表性论著（包括论文与专著，合计5项以内）；②论著之外的代表性研究成果和学术奖励（合计10项以内）。

一、代表性论著

(1) Xiaoying Gan^{(#)(*)}; Yuqing Li; Weiwei Wang; Luoyi Fu; Xinbing Wang, [Social Crowdsourcing to Friends: An Incentive Mechanism for Multi-Resource Sharing](#)[✓], IEEE Journal on Selected Areas in Communications, 2017.3, 35(3): 795~808
(期刊论文)

(2) Xiaoying Gan^{(#)(*)}; Xiong Wang; Wenhao Niu; Gai Hang; Xiaohua Tian; Xinbing Wang; Jun Xu, [Incentivize Multi-Class Crowd Labeling Under Budget Constraint](#)[✓], IEEE Journal on Selected Areas in Communications, 2017.4, 35(4): 893~905
(期刊论文)

(3) Xiaoying Gan^(#); Zesen Zhang; Yiluo Fu; Xinbing Wang, Unraveling Impact of Critical Sensing Range on Mobile Camera Sensor Networks, IEEE Transactions on Mobile Computing
(期刊论文)

(4) Xiong Wang; Riheng Jia; Xiaohua Tian; Xiaoying Gan^(*), Dynamic Task Assignment in Crowdsensing with Location Awareness and Location Diversity, IEEE INFOCOM, 2018.4
(会议论文)

(5) Xiong Wang; Riheng Jia; Xiaohua Tian; Xiaoying Gan^(*); Yiluo Fu; Xinbing Wang, Location-Aware Crowdsensing: Dynamic Task Assignment and Truth Inference, IEEE Transactions on Mobile Computing, 2018.11, 1(1): 1~1
(期刊论文)

二、论著之外的代表性研究成果和学术奖励

(1) 甘小莺(1/8), 基于云WiFi的无线城市运营平台, 上海市人民政府, 科技进步, 其他, 2017.11.16
(甘小莺; 洪峰; 欧丽君; 赵子龙; 易爱冬; 夏丽芳; 树亚; 孙海洋) (科研奖励)



除非特殊说明，请勿删除或改动简历模板中蓝色字体的标题及相应说明文字

参与者 简历

张伟楠，上海交通大学，计算机科学与工程系，讲师，特别副研究员，博导

教育经历（从大学本科开始，按时间倒序排序；请列出攻读研究生学位阶段导师姓名）：

2012/09-2016/06，伦敦大学学院（University College London），计算机科学与工程系，博士，导师：Jun Wang

2011/09-2012/08，上海交通大学，计算机科学与工程系，研究生（一年级后公派出国），导师：俞勇

2007/09-2011/06，上海交通大学，计算机科学与工程系，本科

科研与学术工作经历（按时间倒序排序；如为在站博士后研究人员或曾有博士后研究经历，请列出合作导师姓名）：

1. 2016/08-至今，上海交通大学，计算机科学与工程系，讲师，特别副研究员

曾使用其他证件信息（应使用唯一身份证件申请项目，曾经使用其他身份证件作为申请人或主要参与者获得过项目资助的，应当在此列明）

无。

主持或参加科研项目（课题）情况（按时间倒序排序）：

1. 国家自然科学基金青年项目，61702327，基于多域离散数据的深度学习，2018/01-2020/12，26万元，在研，主持。
2. 国家自然科学基金面上项目，61772333，基于深度神经网络的个性化推荐系统，2018/01-2021/12，64万元，在研，参与。
3. 国家自然科学基金面上项目，81771937，基于深度学习和迁移学习的非结构化临床文本挖掘的方法探索，2018/01-2021/12，64万元，在研，参与。
4. 上海市科委英才扬帆计划，17YF1428200，基于多域类别型数据的深度学习预测模型，2017/05-2020/04，20万元，在研，主持。



代表性研究成果和学术奖励情况

(请注意：①投稿阶段的论文不要列出；②对期刊论文：应按照论文发表时作者顺序列出全部作者姓名、论文题目、期刊名称、发表年代、卷（期）及起止页码（摘要论文请加以说明）；③对会议论文：应按照论文发表时作者顺序列出全部作者姓名、论文题目、会议名称(或会议论文集名称及起止页码)、会议地址、会议时间；④应在论文作者姓名后注明第一/通讯作者情况：所有共同第一作者均加注上标“#”字样，通讯作者及共同通讯作者均加注上标“*”字样，唯一第一作者且非通讯作者无需加注；⑤所有代表性研究成果和学术奖励中本人姓名加粗显示。)

按照以下顺序列出：

一、代表性论著（包括论文与专著，合计5项以内）；

- [1] Cai H., Chen T., **Zhang W.***, Yu Y., Wang J., [Efficient Architecture Search by Network Transformation](#), **32nd AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018.2.2-2018.2.7 (CCF A类)**
- [2] Cao X., Chen H., Wang X., **Zhang W.***, Yu Y., [Neural Link Prediction over Aligned Networks](#), **32nd AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018.2.2-2018.2.7 (CCF A类)**
- [3] Yin L., Han J., **Zhang W.***, Yu Y., [Aggregating Crowd Wisdoms with Label-aware Autoencoders](#), **Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017.8.19-2017.8.25 (CCF A类)**
- [4] Wang X., Yu L.#, Ren K., Tao G., **Zhang W.***, Yong Yu; Jun Wang, [Dynamic Attention Deep Model for Article Recommendation by Learning Human Editors' Demonstration](#), **Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, Canada, 2017.8.13-2017.8.17 (CCF A类)**
- [5] Wang J.*, Yu L., **Zhang W.***, Gong Y., Xu Y., Wang B., Zhang P., Zhang D., [IRGAN: A Minimax Game for Unifying Generative and Discriminative Information Retrieval Models](#), **Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Tokyo, Japan, 2017.8.7-2017.8.11. (CCF A类)**



二、论著之外的代表性研究成果和学术奖励（合计10项以内）。

无。



附件信息

序号	附件名称	备注	附件类型
1	Incentivize Multi-Class Crowd Labeling Under Budge		代表性论著
2	Social Crowdsourcing to Friends: An Incentive Mech		代表性论著
3	Dynamic Task Assignment in Crowdsensing with Locat		代表性论著
4	Location-Aware Crowdsensing: Dynamic Task Assignme		代表性论著



项目名称： 群智数据融合网络关键技术研究

资助类型： 面上项目

申请代码： F020709. 新型感知计算及网络

国家自然科学基金项目申请人和参与者公正性承诺书

本人**在此郑重承诺**：严格遵守中共中央办公厅、国务院办公厅《关于进一步加强科研诚信建设的若干意见》规定，所申报材料和相关内容真实有效，不存在违背科研诚信要求的行为；在国家自然科学基金项目申请、评审和执行全过程中，恪守职业规范和科学道德，遵守评审规则和工作纪律，杜绝以下行为：

- （一）抄袭、剽窃他人科研成果或者伪造、篡改研究数据、研究结论；
- （二）购买、代写、代投论文，虚构同行评议专家及评议意见；
- （三）违反论文署名规范，擅自标注或虚假标注获得科技计划等资助；
- （四）购买、代写申请书；弄虚作假，骗取科技计划项目、科研经费以及奖励、荣誉等；
- （五）在项目申请书中以高指标通过评审，在项目计划书中故意篡改降低相应指标；
- （六）以任何形式打听尚未公布的评审专家名单及其他评审过程中的保密信息；

（七）本人或委托他人通过各种方式及各种途径联系有关专家进行请托、游说，违规到评审会议驻地游说评审专家和工作人员、询问评审或尚未正式向社会公布的信息等干扰评审或可能影响评审公正性的活动；

（八）向评审工作人员、评审专家等提供任何形式的礼品、礼金、有价证券、支付凭证、商业预付卡、电子红包，或提供宴请、旅游、娱乐健身等任何可能影响评审公正性的活动；

- （九）其他违反财经纪律和相关管理规定的行为。

如违背上述承诺，本人愿接受国家自然科学基金委员会和相关部门做出的各项处理决定，包括但不限于撤销科学基金资助项目，追回项目资助经费，向社会通报违规情况，取消一定期限国家自然科学基金项目申请资格，记入科研诚信严重失信行为数据库以及接受相应的党纪政纪处理等。

编号	姓名 / 工作单位名称（应与加盖公章一致） / 证件号码 / 每年工作时间（月）	签字
1	甘小莺 / 上海交通大学 / 3*****1 / 8	
2	张伟楠 / 上海交通大学 / 5*****7 / 6	
3	范桂云 / 上海交通大学 / 3*****4 / 10	
4	秦志达 / 上海交通大学 / 1*****3 / 10	
5	王雄 / 上海交通大学 / 4*****X / 10	
6	李雨晴 / 上海交通大学 / 3*****6 / 10	
7	王彪 / 上海交通大学 / 3*****9 / 9	
8		
9		
10		



项目名称： 群智数据融合网络关键技术研究

资助类型： 面上项目

申请代码： F020709. 新型感知计算及网络

国家自然科学基金项目申请单位公正性承诺书

本单位依据国家自然科学基金项目指南的要求，严格履行法人负责制，**在此郑重承诺：**本单位已就所申请材料内容的真实性和完整性进行审核，不存在违背中共中央办公厅、国务院办公厅《关于进一步加强科研诚信建设的若干意见》规定和其他科研诚信要求的行为，申请材料符合《中华人民共和国保守国家秘密法》和《科学技术保密规定》等相关法律法规，在项目申请和评审活动全过程中，遵守有关评审规则和工作纪律，杜绝以下行为：

（一）采取贿赂或变相贿赂、造假、剽窃、故意重复申报等不正当手段获取国家自然科学基金项目申请资格；

（二）以任何形式探听未公开的项目评审信息、评审专家信息及其他评审过程中的保密信息，干扰评审专家的评审工作；

（三）组织或协助项目团队向评审工作人员、评审专家等提供任何形式的礼品、礼金、有价证券、支付凭证、商业预付卡、电子红包等；宴请评审组织者、评审专家，或向评审组织者、评审专家提供旅游、娱乐健身等任何可能影响科学基金评审公正性的活动；

（四）包庇、纵容项目团队虚假申报项目，甚至骗取国家自然科学基金项目；

（五）包庇、纵容项目团队，甚至帮助项目团队采取“打招呼”等方式，影响科学基金项目评审的公正性；

（六）在申请书中以高指标通过评审，在计划书中故意篡改降低相应指标；

（七）其他违反财经纪律和相关管理规定的行为。

如违背上述承诺，本单位愿接受国家自然科学基金委员会和相关部门做出的各项处理决定，包括但不限于停拨或核减经费，追回项目经费，取消一定期限国家自然科学基金项目申请资格，记入科研诚信严重失信行为数据库以及主要责任人接受相应党纪政纪处理等。

依托单位公章：

日期： 年 月 日

合作研究单位公章：

日期： 年 月 日

合作研究单位公章：

日期： 年 月 日