



Survey TPAMI'24

A Survey on Information Bottleneck

Shizhe Hu (胡世哲), Zhengzheng Lou, Xiaoqiang Yan, Yangdong Ye (叶阳东)*

摘要

该综述论文是为了纪念信息瓶颈理论的创始人之一Naftali Tishby 教授，他于2021年8月去世，享年 68 岁。信息瓶颈 (Information Bottleneck, IB) 是一种用于模式分析和表示学习的新型信息理论方法，自1999年诞生以来受到了广泛的关注。它巧妙地平衡了数据压缩和信息保存，并相应地提高了预测或表示能力。该综述总结了过去20多年来IB的理论进展和实际应用，系统地探讨了IB的基础理论、优化、扩展模型和面向任务的算法。现有的IB方法大致分为两部分：传统IB和深度IB。

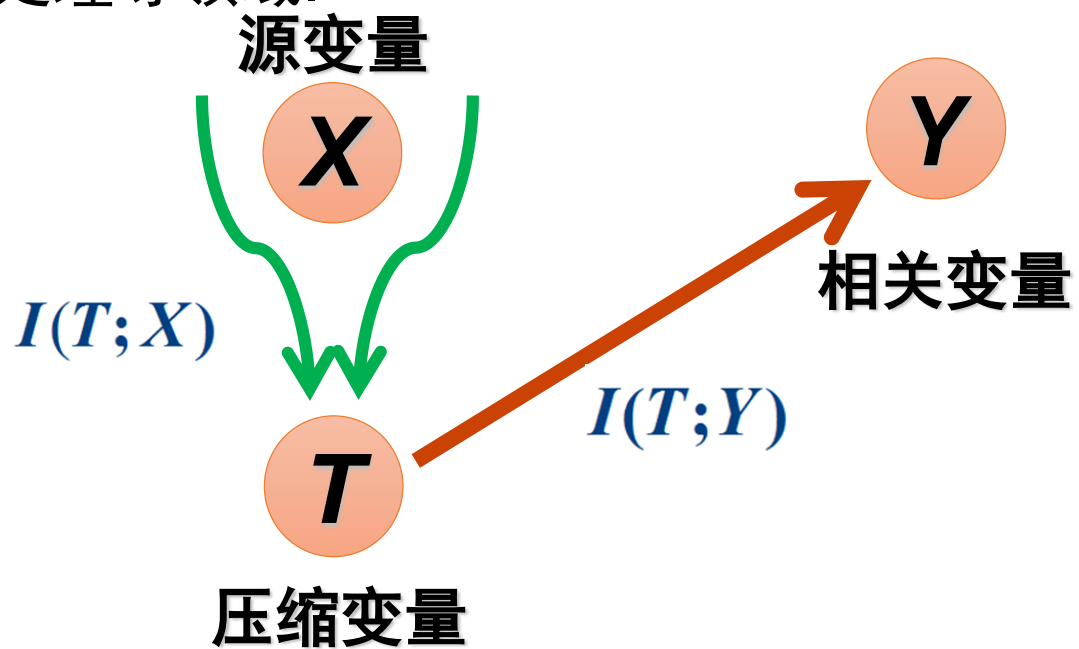
前者包含传统机器学习分析技术优化的且不涉及任何神经网络的IB，后者包括涉及深度神经网络DNN 的解释、优化和改进的IB。具体来说，根据技术分类，**传统IB进一步分为三类**：基础IB、信息化IB和传播IB。根据问题设置分类，**深度IB包括**：辩论：使用 IB 理解DNN、使用IB优化DNN和基于DNN的IB方法。

此外，还讨论了一些值得未来研究的潜在问题。该综述试图描绘出一幅更完整的IB图景，后续的研究工作可从中受益。(链接：<https://ieeexplore.ieee.org/document/10438074>)

信息瓶颈（IB）理论

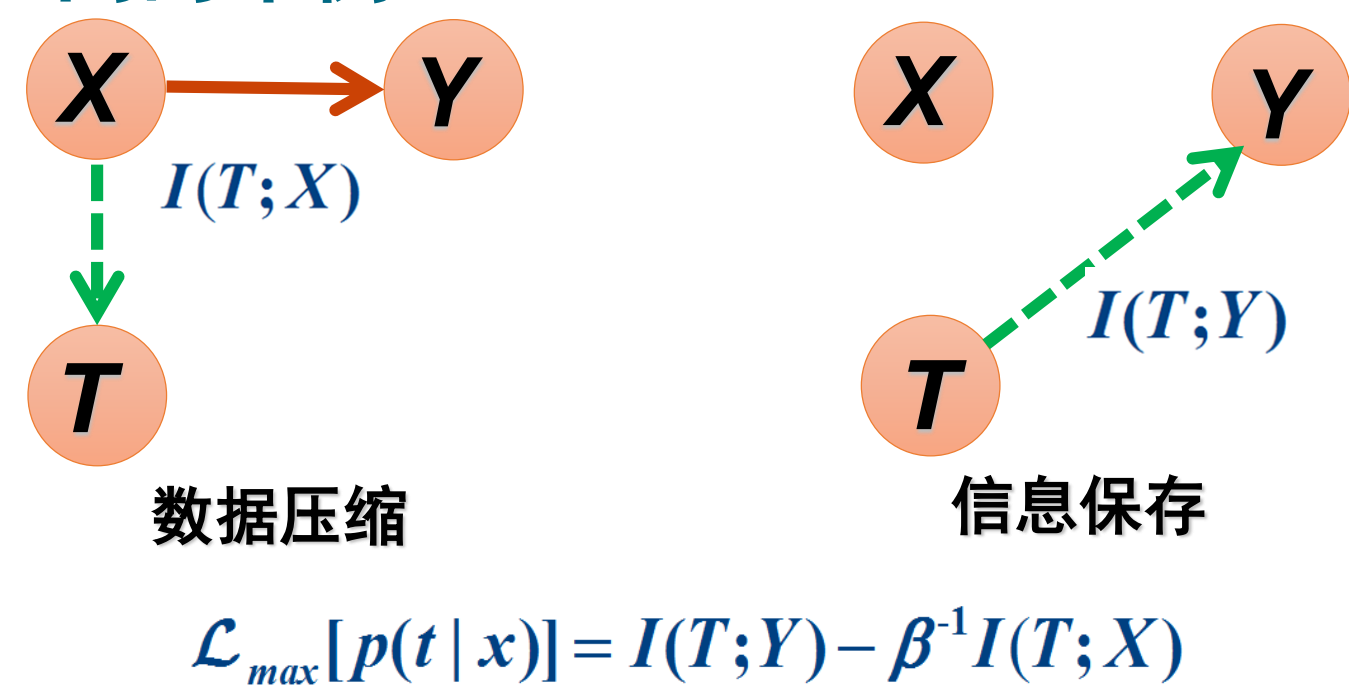
IB (Information Bottleneck)^[1]理论: 基于信息论的数据分析方法，用随机变量X, Y和T来描述领域问题。

- (1) 良好的信息理论基础。
- (2) 已成功地应用在文档数据分析，图像/视频识别，语音识别，自然语言处理等领域。



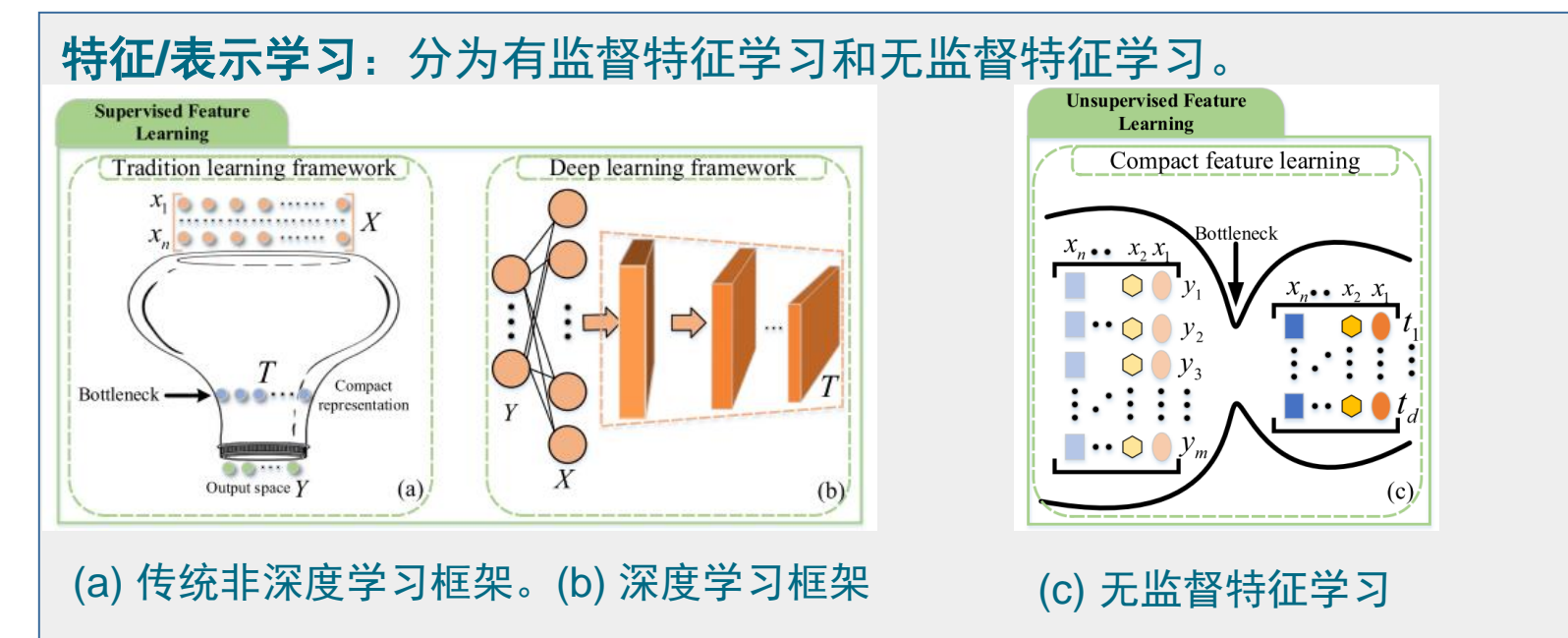
[1] Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. In Proc. Annual Allerton Conf. Communication Control Computing, 368–377, 1999.

IB形式化表示：注重数据压缩与信息保存之间的平衡



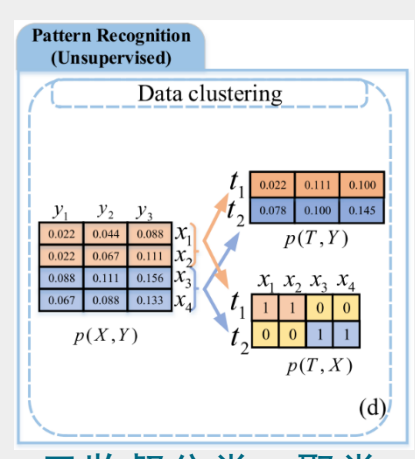
IB可解决的科学问题

- **特征/表示学习**：学习原始数据好的特征表达，以用于下游任务。机器学习中，同样数据的不同表达会直接决定下游任务的难易程度，一个好的数据表示常常是机器学习的核心任务。
- **模式识别**：识别数据的潜在模式结构，将样本划分到一定的类别中。模式识别以图像处理与计算机视觉、语音语言信息处理等为主要研究方向，研究人类模式识别的机理以及有效的计算方法。



模式识别：分为有监督分类和无监督分类。

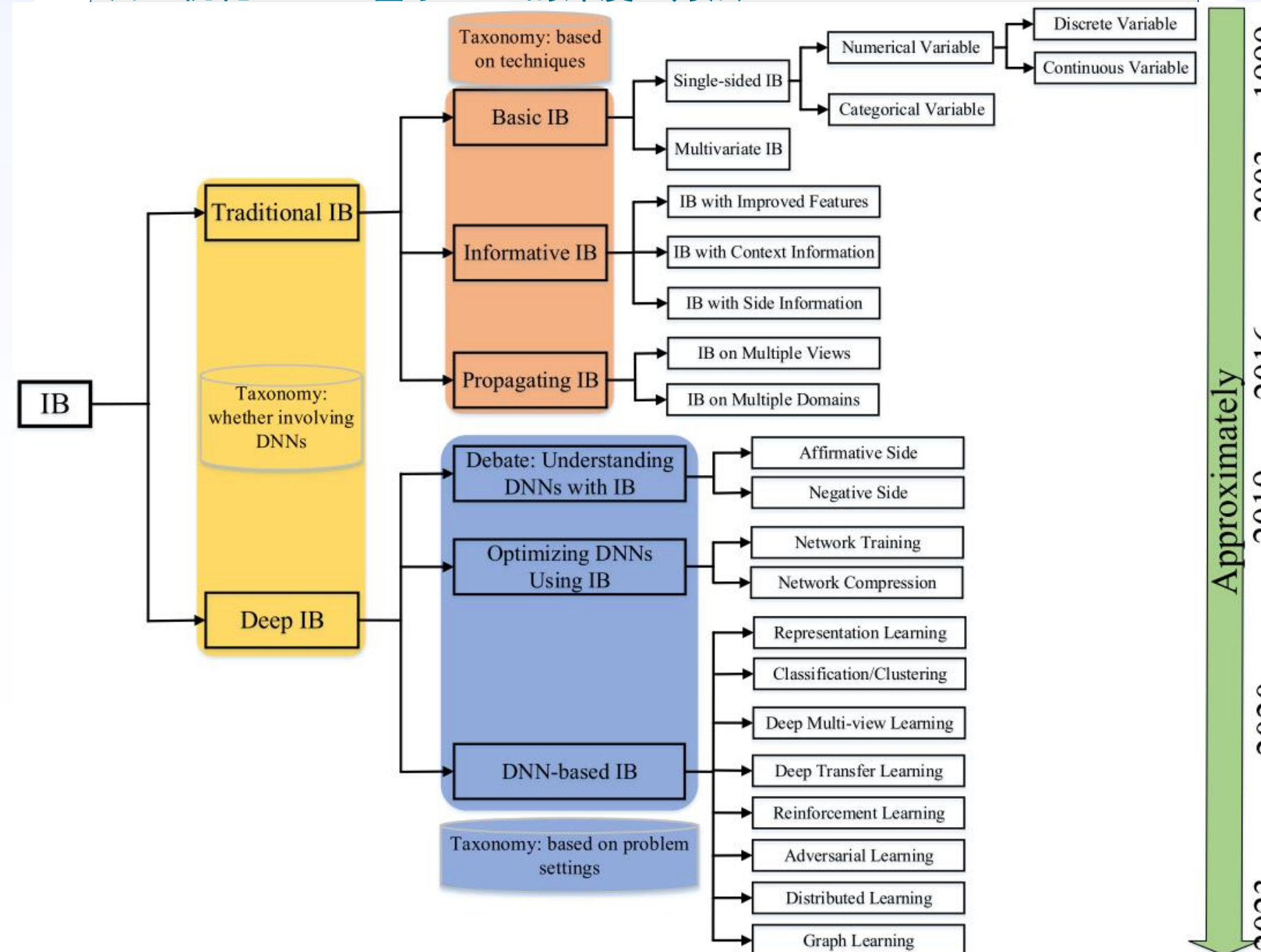
有监督分类：
有监督特征表示学习+分类器(SoftMax, SVM)。



无监督分类：聚类

IB方法国内外研究分类：

- (1) 按照是否涉及DNN，分为**传统IB**和**深度IB**。
- (2) 基于方法学，**传统IB**分类三类：基础IB、信息化IB、传播IB。
- (3) 基于问题设置，**深度IB**分为三类：关于IB理解DNN的争论、使用IB优化DNN、基于DNN的深度IB方法。



IB背后的哲学道理

哲学道理1：有效的总是简单的。

鉴于各种复杂的输入数据，IB试图通过保留最具鉴别性的信息来提取复杂输入的重要表征，同时消除噪声和冗余信息，从而使其更为简单。

尽管这些鉴别性信息简单，但在许多领域非常有效，尤其是特征表示学习和模式识别。

哲学道理2：影响我们理解世界的不是多数特性而恰恰是少数。

数据的数以万计的特性/属性并不能很好地帮助我们理解世界。比如，我们常常记忆或者识别某个物品或者某个地方仅仅靠一些关键的特性或属性。

同样地，IB尽可能将数据压缩至很小的、关键的一部分，却很紧凑、有鉴别性

哲学道理3：最高效的学习不是记忆知识而是压缩模式提取。

回想一下，学生在年轻的时候通常通过死记硬背来学习新事物或能力。随着时间的流逝，他们会忘记大多数。

相比之下，高效的学习是抽象底层的模式、概念或规则，然后逐渐吸收它们。

同样，IB学习输入数据的紧凑和相关模式，同时“忘记”（即丢弃）不相关的部分数据，从而提高了IB的泛化能力。

我们对于IB理论的贡献

国家自然科学基金面上项目 ：IB算法及其应用的研究，2008-2010，主持：叶阳东 问题 ：IB算法解精确度、簇数自动划分 场景 ：图像压缩、信息编码、信息安全 成果 ：ACM MM, PRICAI, PR Letters、电子学报、模式识别与人工智能、铁道学报	国家自然科学基金面上项目 ：多变量IB方法及算法的研究，2012-2015，主持：叶阳东 问题 ：协作模型构建，平衡参数确定 场景 ：信息安全、医学诊断、图像模式分析 成果 ：ACM MM, IJCAI, KBS, PR Letters, CIKM, PRICAI, 计算机学报、计算机研究与发展、电子学报、模式识别与人工智能
国家自然科学基金面上项目 ：传播IB方法的研究，2018-2021，主持：叶阳东 问题 ：模型建立、传播机制构建、参数学习 场景 ：隐藏信息分析、多传感器监控、信息推荐 成果 ：AAAI, CVPR, ICASSP, IEEE TMM, IEEE TCYB, IEEE TKDE, ACM TKDD, Information Fusion, 计算机研究与发展、软件学报	国家自然科学基金面上项目 ：图传播IB方法的模型及传播机制研究，2022-2025，主持：叶阳东 问题 ：模型构建、图传播机制研究 场景 ：人群计数、微视频场地估计、物品推荐 成果 ：AAAI, CVPR, AAAI, ACMMM, IEEE TPAMI, IEEE TCYB, IEEE TNNLS, IEEE TMM, IEEE TSMC, 中国科学:信息科学等

主要完成人简介

叶阳东 (最新动态可访问<http://www5.zzu.edu.cn/mlis/>)，计算机与人工智能学院二级教授，博士生导师。曾在澳大利亚迪肯大学做高级访问学者1年，长期从事机器学习、知识工程、智能系统、智能电网等方面的研究。主持国家重点研发计划子课题1项、国家自然科学基金4项、河南省科技攻关计划项目1项，参与国家自然科学基金9项、863项目1项。在IEEE汇刊、CVPR、AAAI、ACM Multimedia、IJCAI、中国科学：信息科学、计算机学报、软件学报等国内外知名期刊与会议发表论文100余篇，获批国家发明专利10余件，成果获河南省科技进步三等奖。主讲数据库原理、计算科学导论和Petri网理论等本科/研究生课程，参与《数据库原理教程》、《计算机引论》等教材编著，培养博/硕士研究生110余人。兼职包括中国计算机基础教育研究会理事、河南省计算机学会理事等。

胡世哲 (最新动态可访问<https://shizhehu.github.io/>)，计算机与人工智能学院副研究员，硕士生导师，中国图象图形学学会视觉大数据专业委员会委员，IEEE/ACM/CCF/CSIG会员。主要研究方向有信息瓶颈(Information Bottleneck)理论、多模态学习、多视图学习、可信/可解释学习、聚类分析等。已以第一/通讯作者身份在IEEE TPAMI、TIP(x2)、TKDE、TNNLS(x2)、TCYB、Pattern Recognition、ACM MM、IEEE ICASSP、计算机学报等中国计算机学会A/B类会议期刊或中科院1区期刊上发表SCI/EI论文10余篇。论文发表后得到国内外同行的关注和积极评价，Google学术被引用近500次。多次担任国际顶级期刊IEEE TIP、TKDE、TNNLS、TMM、Information Fusion、Neural Networks以及国内外顶级会议CVPR、AAAI、CCDM、PRCV、PRAI的审稿人。主持国家自然科学基金青年基金项目、中国博士后基金面上项目各1项，参与国家重点研发计划专项、国家自然科学基金面上项目等3项。曾获ACM中国郑州分会优秀博士学位论文奖、河南省自然科学学术二等奖、郑州大学优秀博士学位论文。

致谢

- [1] 所有合作者
- [2] 信息瓶颈理论驱动的自监督多视角聚类方法研究，62206254，2023-2025，国自然青年基金项目，主持：胡世哲
- [3] 图传播IB方法的模型及传播机制研究，62176239，2022-2025，国家自然科学基金面上项目，主持：叶阳东
- [4] 传播IB方法的研究，61772475，2018-2021，国家自然科学基金面上项目，主持：叶阳东