**异质信息网络的研究现状和未来发展**

**1.背景**

（1）现存的系统

现实生活中的大多数系统，是由大量相互作用、不同类型的组件构成的。为了更好地进行分析，通常将其建模为同质信息网络 (homogeneous information network)。

（2）存在的问题

这种建模方法，往往只抽取了实际交互系统的部分信息，或者没有区分交互系统中对象及关系的差异性，这样往往会造成信息的不完整或信息损失。

（3）改进的技术

将这些互连的多类型网络化数据建模为异质信息网络 (heterogeneous information network)，并且通过利用网络中丰富的对象和关系信息来设计结构分析方法。

（4）优势

与广泛研究的同质信息网络相比，异质信息网络包含全面的结构信息和丰富的语义信息。

**2.概念介绍**

异质信息网络：被定义为一个有向图。它包含多种类型的对象或者关系，每个对象属于一个特定的对象类型，每个关系属于一个特定的关系类型。

网络模式：是定义在对象类型和关系类型上的一个有向图，是信息网络的元描述。

**3.异质信息网络**

（1）应用现状

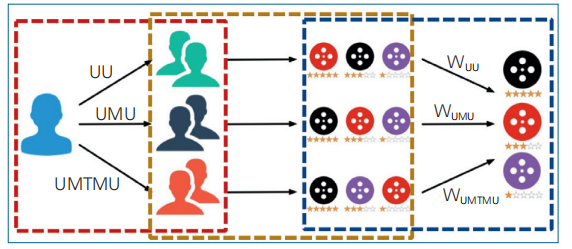
异质信息网络已经广泛应用于主要的数据挖掘问题，特别是相似性度量 、聚类 、分类 、链接预测 、推荐等任务。

（2）经典应用

基于异质信息网络的语义推荐

石川等人提出了个性化语义推荐方法 SemRec，

作为一种基本的推荐技术，协同过滤方法通过相似的用户对用户进行推荐。在异质网络中，可以利用元路径找到不同特性的相似用户。例如，通过元路 径“UU”（U、M、T 分别表示用户、电影、电影类型），可以找到用户的朋友，这对应于社会化推荐 ；通过元路径“UMU”，可以找到具有相同观影记录的用户， 这对应于传统的协同过滤。不同的相似用户有不同的推荐结果，有效整合这些推荐结果，可以产生综合的最终推荐。该方法还考虑了用户和电影之间打分关系上的分值（即关系权重），提出了带权异质信息网络和带权元路径等概念，以及相应的相似性计算方法。此外，该方法还采用组推荐技术对具有相同打分偏好的用户进行聚类。实验表明，由于融合了更多信息，该方法不仅具有更高的推荐准确性，而且能够有效缓解冷启动问题。此外，该方法能够根据用户的打分特性对用户进行聚类，较好地反映了用户群体特征。



基于元路径的语义推荐方法

**4.未来发展**

（1）更加复杂的网络构建

实际应用中，用真实数据构造异质信息网络会遇到很多挑战。对于关系数据库之类的结构化数据，构造异质信息网络比较容易，然而即使是在这种网络中，对象和关系也可能具有噪声，会出现对象重名或关系不完整等问题 ；对于文本、图像等非结构化数据，如何准确抽取出相应的对象和关系，进而建立更加完善和准确的异质信息网络，也将面临更多挑战。

（2）更加强大的分析方法

异质信息网络中的对象和关系包含着丰富的语义信息，而元路径可以捕捉这种语义信息。异质信息网络上的很多数据挖掘任务是基于元路径进行研究的，但是元路径在某些应用场景中并不能捕捉到精细的语义信息。

对更加复杂的网络结构（如知识图谱），如何设计更加灵活精细的语义探索工具仍然需要进一步的研究。

（3）更大数据的处理

多样性是大数据的重要特征，异质网络是处理大数据多样性的有效方法。然而，构建一个真正的基于异质网络的大数据分析系统也是具有挑战性的工作。实际上，异质网络是巨大的，甚至是动态的，所以通常不能在内存中直接进行处理。由于用户往往只对一小部分节点、链接或子网络感兴趣，我们可以根据用户需求，从现有网络中动态地提取子网络进行分析。另外，设计基于异质网络的快速算法和并行算法也是亟需研究的内容。