科技情报中多源信息融合的模式构建

马红岩 陈 峰 曾 文
（中国科学技术信息研究所，北京 100038）

摘要：首先综述科技情报研究中多源信息融合的主要数据类型和特点；然后按照融合深度的不同，将多源信息融合划分为特征及类型融合、关系融合和聚类融合3种类型，并分析每一类型中多源信息融合的相关研究现状；最后以科技情报领域主题识别问题为例，从特征及类型融合、关系融合、聚类融合3个层面提出并构建多源信息融合的新模式。

关键词：多源信息融合；关系融合；聚类融合；科技情报；融合模式
DOI：10.3772/j.issn.1674−1544.2022.03.001
CSTR：15994.14.issn.1674−1544.2022.03.001
中图分类号：G353.1
文献标识码：A

Model Construction of Multi-source Information Fusion in Science and Technology Information

MA Hongyan, CHEN Feng, ZENG Wen
(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: This paper analyzes the main data types and characteristics of multi-source information fusion in scientific and techical information research, and divides multi-source information fusion into three types: feature and type fusion, relationship fusion and cluster fusion according to different fusion depths, and systematically summarizes the related research status of multi-source information fusion in each type. On this basis, taking the subject recognition in the field of scientific and techical information as an example, this paper puts forward and constructs a new mode of multi-source information fusion from three aspects: feature and type fusion, relationship fusion and cluster fusion.

Keywords: multi-source information fusion, relationship fusion, cluster fusion, scientific and technical information, fusion model

0 引言

随着大数据时代的到来，各种类型的信息呈现爆炸似的增长，如何在海量的信息中对不同类型、不同来源的信息进行综合分析变得尤为重要。在科技情报领域，情报检测、热点发现、前沿识别、科技评价、科技查新等都是科技情报工作和科技情报研究的主要任务，而这些任务的完成都需要多种来源的信息[1]，即多源信息。对这些多源信息进行综合进而生成有效信息的过程，即为多源信息融合。多源信息融合（又称信息融合、数据融合、多传感器信息融合）起源于 20 世纪 70 年代的军事领域，最初目的是美国为解决军事需求，将指挥（Command）、控制（Control）、通

作者简介：马红岩（1997—），女，中国科学技术信息研究所硕士生，研究方向为技术竞争情报；**陈峰**（1965—），男，中国科学技术信息研究所研究员，研究方向为竞争情报、技术预见、科技政策与发展战略等；**曾文**（1973—），女，中国科学技术信息研究所研究员，研究方向为科技情报分析理论与方法、区域创新发展研究等（通信作者）。

基金项目：国家自然科学基金面上项目“基于开源情报的科技前沿多维度探测方法及模型研究”（72074201）。
收稿时间：2022 年 1 月 7 日。

信（Communication）和情报（Intelligence）（即C3I）军事系统中的数据进行多源相关性融合，并将其作为国防重点开发项目，此后迅速发展成为一门独立学科[2]。随后广泛应用于各个领域中，而每个领域对其的定义又各不相同。笔者认为，在科技情报领域中，信息融合是指对多源异构的科技数据或科技信息进行整合，以获得更加完整、准确、可靠科技情报的一个过程。在科技情报分析中利用单一信息源提供情报支持显然存在信息不够全面客观、情报具有片面性、对科技决策的支持力度不足的问题。多源信息融合是解决这一问题的主要途径之一。因此，利用多种信息源的融合辅助科技情报的生产是必要的。

鉴于此，本文将在调研科技情报领域中多源信息融合相关研究的基础上，梳理多源信息融合的主要数据类型和特点，分析和总结科技情报研究中多源信息融合的主要数据类型，并结合计算机技术，如机器学习，提出面向科技情报领域主题识别的多源信息融合新模式。

1　多源信息融合的主要数据类型和特点

本文对科技情报研究中多源信息融合的主要数据类型及特点[3-4]进行总结（表1），为进一步提出和构建多源信息融合新模式提供支撑。

2　多源信息融合的分类

国内外学者对多源信息融合展开了大量的研究。在国内，化柏林[5]首先提出在情报分析中要引进融合论，并强调将其作为情报研究工作中的主要方法论，随后他从多源信息类型的角度出发，将多源信息划分为同型异源信息、异质异构信息以及多语种信息，并分析了不同类型多源信息的融合方法[1]。除此之外，化柏林等[6]还系统阐述了大数据环境下多源信息融合的理论基础、融合技术和方法，探讨了大数据背景下多源信息融合的应用。在国外，Avila等[7]融合专利和论文

表1　多源信息融合的主要数据类型和特点[3-4]

序号	数据类型		特点
1	专利		专利文献是技术信息的主要载体，是世界上获取技术信息的最大来源。实证证明，全球90%以上的技术信息来源于专利文献，包含了丰富的技术信息、法律信息和经济信息，其所介绍的技术在内容上具有新颖性、创造性和实用性，是科技情报工作及学术研究中的重要数据类型之一。专利文献中传递的技术情报可以帮助我们及时了解世界科技动态
2	论文	期刊论文	期刊论文是发表在杂志上的学术论文，出版周期短，能及时反映国内外各学科领域的新动向、新进展、新技术以及新成果，数量庞大且针对性较强，也是科技情报工作及学术研究中的重要数据类型之一
		学位论文	学位论文包括学士论文、硕士论文和博士论文。其中，博士论文具有较大独创性，且质量最高，其研究的问题较为深入，专业性较强，是情报研究中重要的情报源之一
		会议论文	会议论文是在学术会议中发表的论文，一般以论文集的形式进行发表能够及时反映各专业各学科的新成果，内容上具有新颖性。会议论文一般都是定期发表的，针对性较强
3	基金项目		基金项目是由各国各地区为鼓励科技创新而设立的，通过基金项目可以及时了解各个国家及地区的科技动态，反映了研究的最新动向，甚至从项目信息中可以体现国家的科技计划与发展战略，具有“将来时”的特性。其不足是主题较为宏观且不够丰富，粒度较大
4	科技报告		科技报告是记录科研活动各个阶段中科学、技术研究结果或研究进展的一种特种文献，其时效性高，能够迅速及时反映新的研究成果，且内容广泛、完整、具体、技术含量高
5	图书		图书是用来记录一切成果、成就的主要载体，所承载的内容较为全面、系统，相较于其他出版物，其出版周期较长，传递信息不够及时，前瞻性较弱
6	科技规划文本		科技规划文本是科技政策的体现形式，其所包含的主题目前正处于研究的状态中，还未取得突破性的成果，与其他文献相比，其所包含的信息具有更多的前瞻性，是当前情报研究的重要情报源。不足是主题较为宏观，粒度较大
7	网络舆情		网络舆情信息是公众对技术发展最直接的认知反馈
8	技术标准		技术标准的制定是以科技研发及其相关科技成果为基础的，技术标准是技术传播有力的载体
9	会议信息		会议信息往往可以反映出学科领域的最新进展以及最新前沿

两种数据的直接引用、共被引、文献耦合等 3 种关系评估新兴技术的知识建设动态。Wang等[8]提出融合自由贸易区（FTZ）平台共享数据、互联网新闻文本数据和互联网统计数据 3 种信息的融合方法。本文按照融合深度不同，将多源信息融合划分为特征及类型融合、关系融合和聚类融合 3 种类型，并分别对 3 种类型的多源信息融合方法及相关研究进行梳理分析。

2.1 特征及类型融合

2.1.1 特征融合

在特征融合方面，化柏林[1]认为特征融合包括相同特征相同标识、相同特征不同标识、互补型特征、差异型特征等 4 种特征类型的融合。张付志等[9]提出了基于度量级融合的论文特征提取方法来提取PDF论文中Title、Author、Abstract、Keywords等 4 个字段，并将这些特征集成到统一的框架中。本文认为，特征融合是将多源信息的内部特征（如标题、摘要、关键词等）和外部特征（如作者、时间、被引频次等）融合成统一的形式，包含特征匹配、特征拆分、数据滤重、异构加权等 4 个步骤，如图 1 所示。

（1）特征匹配。特征匹配是对相同特征不同标识的数据进行分析与特征映射。同型异源信息其相同特征的表征方式不同，如某一篇论文的标题字段在中国知网（CNKI）被表征为“篇名”，而在万方数据库则被表征为“题名”。对于这种数据首先需要将标识进行统一，如将两种来源的标题字段统一命名为“标题”，然后采用线性融合的方式进行融合。

（2）特征拆分。特征拆分包括多值同特征和多值异特征两种情况。多值同特征，如作者、机构、关键词等特征需要拆分。多值异特征，如德温特数据库中的CP字段既包含引用专利的专利号，又包含引用专利的作者，这样的字段所包含的数据特征是不同的，需要进行拆分。

（3）数据滤重。多种来源的数据相互交叉，重复数据较多，如SCI、EI数据库中收录多种相同的期刊，数据滤重的关键是确定数据的唯一标识，去掉重复数据。如基金项目数据可以使用资助项目号作为唯一标识，期刊论文可以使用DOI作为唯一标识，而对于没有唯一标识的数据，可以采用多种特征相结合的方式作为数据的唯一标识。不同类型的数据其唯一标识不同，需要根据数据的具体情况进行分析。

（4）异构加权。异构加权是对相同特征不同来源的数据采用融合函数进行线性融合，对于融合函数中权重的设置主要有专家打分法和基于统计实证的方法。

2.1.2 类型融合

类型融合是将异质异构信息经过加工处理后融入同一分析目标，涉及特征匹配、特征拆分等处理方法。在数据类型融合方面，Wang等[8]融合自由贸易区（FTZ）平台共享数据、互联网新闻文本数据和互联网统计数据 3 种信息，提出融合过程中的关键点是解决多源异构数据之间的不一致性问题，其在文本主题分类和量化过程中采用朴素贝叶斯多标签分类算法，克服多源数据的结构不一致性，采用时差相关分析法解决两个时间序列的时间不一致性。徐璐璐等[10]将收集到的论文、专利以及基金项目数据进行文本格式转换、分词、去除停用词、字符删除及句法分析等预处理后进行词汇语言的合并，对合并后的文本数据进行主题识别。冯佳等[11]不同数据类型中的特征映射到一个统一的框架上，将多种类型数据统一成相同的形式，该框架的目的是以半结构化的方式表征不同数据类型中的情报信息。

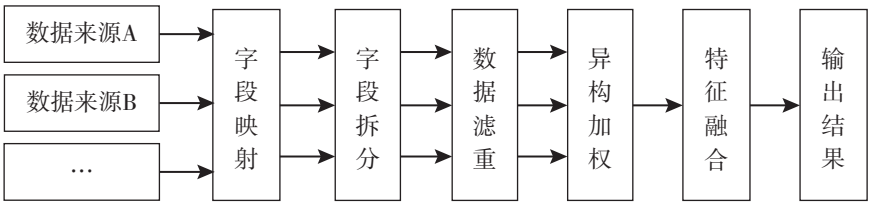

图 1 特征融合示意

2.2　关系融合

关系融合又称多元关系融合，首先获取多种数据的关系包括结构关系（如直接引文、耦合关系、共被引关系、合著关系等）、主题关系（主题词共现）、语义关系等，然后将多种关系融合成一个新的关系以揭示实体之间的关联情况，如图2所示。关系融合的方法主要分为两种：一是将多个距离矩阵进行运算得到关系融合矩阵，二是利用主题模型的方法进行关系融合。

2.2.1　将多个距离矩阵进行运算得到关系融合矩阵

Amjad等[12]、Calero等[13]、滕立[14]、Bei等[15]根据研究场景的不同选取作者－文献矩阵、文献－期刊矩阵、词共现矩阵、词－文献矩阵、参考文献－文献矩阵、作者共现矩阵、作者－机构共现矩阵、作者－国家共现矩阵和机构－国家共现矩阵等适宜的矩阵进行运算形成融合矩阵，分别应用在领域主题分类、科学出版物间知识创造和流动过程分析、总结领域中知识交流模式等应用场景中。Zhang等[16]融合文本内容的语义关系和IPC分类信息的分类关系进行专利组合的混合相似度计算，并应用到技术相似性的研究中，实证表明基于语义关系和分类关系的相似度计算具有一定的可靠性。Avila等[7]构建论文－专利在多关系（直接引用、共被引、文献耦合）呈现下的单一知识网络，用于评估新兴技术的知识建设动态。谭晓等[17]结合多源数据进行知识融合，将多种实体和多种数据关系进行融合，并将其与主题模型进行关联，形成包括结构关系、语义关系和共现关系3种关系融合的网络。

2.2.2　利用主题模型的方法进行关系融合

Tang等[18−19]融合作者、期刊、文档3种信息对LDA主题模型进行改进，提出ACT（Author−conference−topic）模型。Xu等[20−21]和史庆伟等[22]对LDA主题模型进行改进，将主题、作者与时间关联提出作者主题演化模型（Author−Topic over Time，AToT），用于挖掘科技文献中隐含的主题和作者的研究兴趣随时间变化的规律。Du等[23]将专利的作者合作网络和专利－发明人网络融合成一个新的异质网络，并利用该网络对发明人的重要性进行排序。冯佳等[11]从载体－特征－关系3个层面深入全面的构建多源数据融合模型，其中关系融合从内容相关性和逻辑语义性两个方面采用LDA主题模型来实现多元关系的融合。此外，Xu等[24]从融合方式、融合范围、融合深度以及融合数量等多个角度对关系融合方法进行阐述。

2.3　聚类融合

聚类融合又称聚类集成，是对同一类型数据多次聚类的聚类结果或对多种来源多种类型数据的聚类结果进行融合，其中聚类算法可以采用同一聚类算法或不同聚类算法，融合方法主要是利用融合函数对聚类簇进行融合从而得到最终聚类结果[25]。聚类融合如图3所示。具体描述如下：给定包含N个数据类型的数据集$X=\{x_1,x_2,\cdots,x_N\}$，其中第i种数据集$x_i=\{x_{i1},x_{i2},\cdots,x_{iH}\}$，$i$表示第$i$种数据类型$(i=1,2,\cdots,N)$，$H$表示第$i$种数据类型中数据的个数。对数据集X进行$N$次聚类得到$N$个聚类结果的聚类成员集$\beta=\{\beta_1,\beta_2,\cdots,\beta_N\}$，其中第$i$种数据类型的聚类结果$\beta_i=\{c_{i1},c_{i2},\cdots,c_{ik}\}$，$i$=1,2,⋯,$N$，$k$是聚类成员$\beta_i$的聚类簇的数量。聚类融合的目的是设计一种融合函数$\theta$将所有的聚类成员

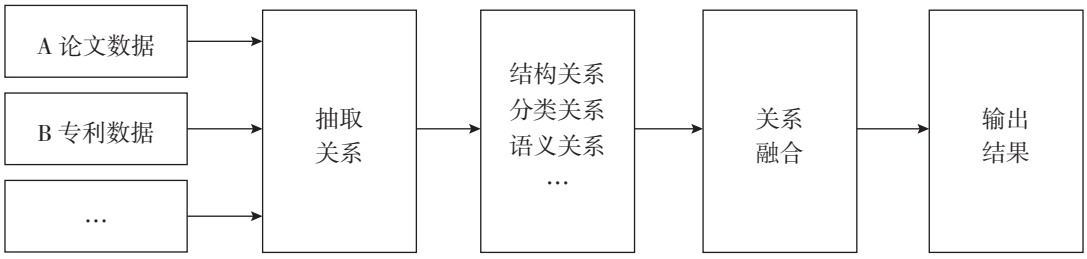

图2　关系融合示意

$\beta_1,\beta_2,\cdots,\beta_N$融合为一个新的聚类结果$\beta^\theta$。

聚类融合过程中的重点是融合函数的设计。融合函数又称一致性函数、共识函数，是指通过对聚类成员进行融合，得到一个统一的聚类结果，是聚类融合过程中的一个重要步骤，融合结果的好坏取决于融合函数的确定。常用的融合函数及代表性研究参见表 2。

3 多源信息融合的新模式

科技情报研究中不同数据类型具有不同的特点，数据类型的选取与研究问题相关。本文针对现有研究方法的局限性，以面向科技情报领域的主题识别问题为例，选取论文（期刊论文和会议论文）、专利、基金项目等 3 种类型的科技数据，

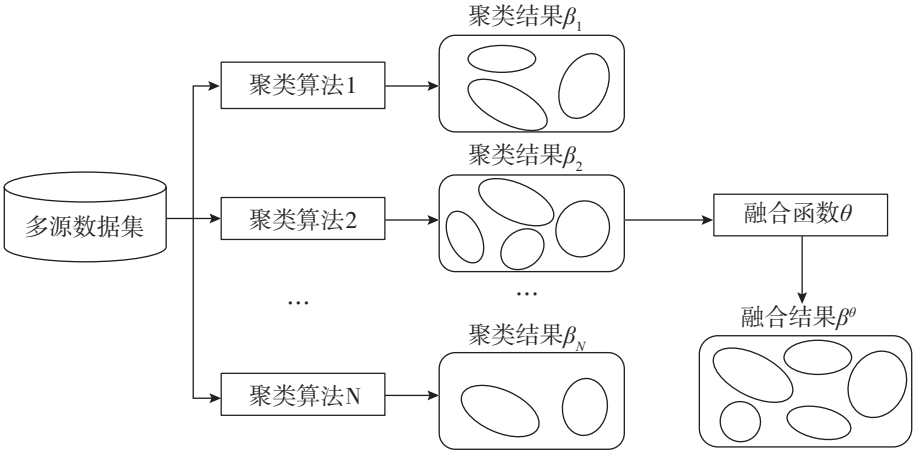

图 3 聚类融合示意

表 2 常用的融合函数及代表性研究

序号	名称	主要思想	代表性研究	优缺点
1	投票方法	在聚类簇标签对齐的基础上，根据数据集合中的数据对象对应的聚类簇进行投票，计算每个数据对象被分到每个聚类簇的次数，数据对象最终划分的簇为划分次数最多的簇[26]	Tumer 等[27] 张静静等[28] 陈晓云等[29]	操作简单，易受聚类成员质量的影响，当聚类成员的质量较差时融合后的结果就会比较差
2	证据累积法	首先将每一个基聚类成员看成一个证据，得到聚类成员的共联矩阵。该矩阵中的元素是数据对象在基聚类成员中同属于一个聚类簇的总次数。然后将聚类融合看成一个新的聚类问题，将共联矩阵作为相似性数据使用聚类算法对数据进行二次聚类从而得到最终的聚类结果[30]	Huang 等[31] Lai 等[32] 毕凯等[33]	该方法使用共联矩阵作为相似矩阵，考虑了数据对象之间的关联，计算较为简单。不足是没有考虑聚类簇之间的关联和其他特征
3	超图划分的方法	首先根据基聚类结果将数据集中的数据对象转换为超图。超图中的点是数据对象，边可以是数据对象之间的相似度值也可以是聚类成员中的聚类簇。然后采用基于图论的聚类算法对超图进行划分得到最终的聚类结果	Strehl 等[25] 王田[34] Huang 等[31]	该方法利用聚类成员表示数据集的结构，充分考虑数据对象和聚类簇之间的关联。不足是利用该方法划分聚类的大小较为相似
4	概率积累的方法	该方法是在证据累积方法的基础上提出的。首先根据基聚类结果使用簇密度计算数据对象之间的距离，得到每个数据对象的component矩阵。然后计算所有component矩阵的平均值，并将其作为p−association矩阵，再对该矩阵采用MST方法进行融合，进而得到最终的融合结果	Wang 等[35]	该方法弥补了证据累积的不足，充分考虑了聚类簇的特征。缺点是计算复杂性较高
5	基于余弦相似度的方法	首先利用余弦相似度算法计算聚类簇之间的相似度，当两个聚类簇之间的相似性达到一个阈值，认为这两个聚类簇的主题相同。然后采用线性融合的方式进行类群之间的融合	许晓阳等[36]	基于余弦相似度的方法考虑了数据对象和聚类簇之间的关联，计算较为简单
6	基于信息熵的方法	首先对每个聚类簇的不稳定性进行衡量，结合信息熵的概念和Jaccard系数提出聚类簇评价指标，从簇层面进行加权得到加权矩阵。然后进行二次聚类，将加权矩阵作为相似性数据采用适合的聚类算法进行划分后得到最终的融合结果	邵长龙等[37]	基于信息熵的方法考虑了每个聚类簇的不稳定性，从簇层面进行融合。缺点是操作较为复杂

从特征及类型融合、关系融合、聚类融合3个层面提出并构建了多源信息融合新模式，如图4所示。

以下是信息融合模式的基本思路及采用的方法。

第一层：特征及类型融合。选择一个发展相对成熟，边界相对清晰的领域作为分析对象，收集该领域多种来源的会议论文、期刊论文、专利、基金项目数据，经过特征匹配、特征拆分等处理后融合到一个统一的框架上。该框架包含内部特征和外部特征两个部分，如图5所示。

第二层：关系融合。根据应用场景的不同获取不同的关系，科技领域主题识别的多源信息融合模式主要融合基于文本内容的语义关系、基于分类信息的分类关系、基于参考文献的引用关系，形成基于多关系融合的文本相似度矩阵，用于后续的文本聚类。

步骤1：构建文本向量获取语义关系。首先抽取文献的标题和摘要，对其进行分词、去标点符号、去停用词等处理；然后选择如Doc2vec模型、Word2vec模型、词袋模型等文本向量化技术构建文本向量。

步骤2：构建分类向量获取分类关系。首先获取文献的分类信息，如EI数据库下载的论文包含EI分类号字段，Web of Science数据资源中下载的论文包含研究方向字段，论文的分类信息可以从EI分类号、研究方向等字段获取；专利数据的分类信息可从IPC分类号字段中获取；基金项目数据目前没有分类信息。然后通过获取的分类信息构建分类向量，向量的元素值是该文献与其他文献的共同分类强度，即两篇文献共同的分类号数量与二者分类号并集中元素数量的比值。两篇文献共同分类号的数量越多，表明这两篇文献的研究主题越相似。分类向量元素值计算公式为：

$$A_{ij}=\frac{B_{ij}}{C_{ij}} \quad (1)$$

其中，i，j分别表示第i篇和第j篇文献，A_{ij}

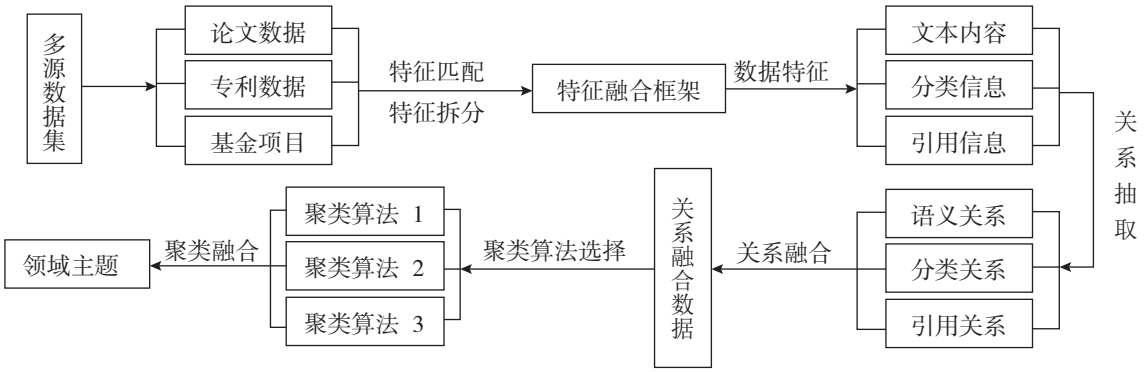

图4　多源信息融合模式

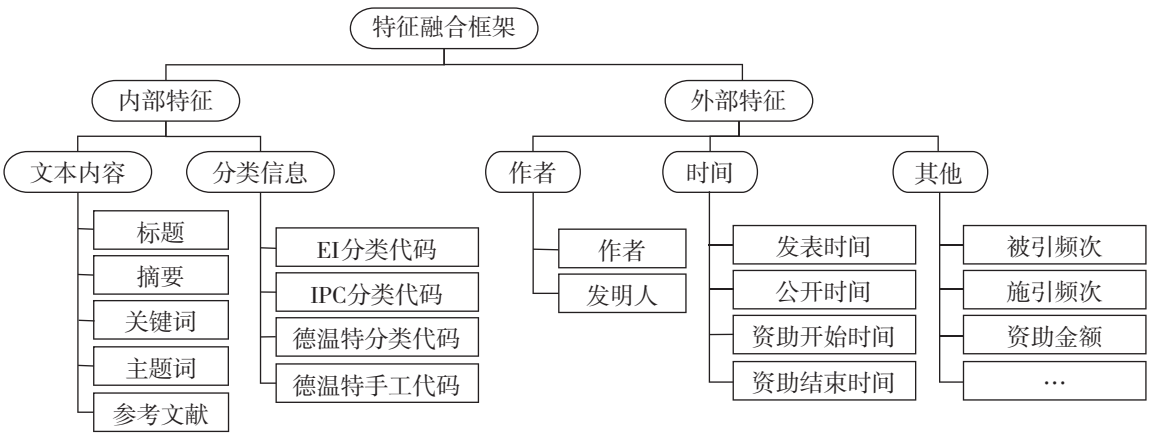

图5　特征融合框架

表示第i篇文献第j个元素值，B_{ij}表示第i篇文献和第j篇文献共有分类号的数量，C_{ij}表示第i篇文献和第j篇文献分类号并集中的数量。

步骤3：构建引用向量获取引用关系。根据文献的参考文献构建引用向量，向量的元素值是该文献与其他文献的共同引文强度值，即两件文献共同的参考文献数量与二者参考文献并集中文献数量的比值。两篇文献共同参考文献的数量越多，表明这两篇文献的研究主题越相似。引用向量元素值的计算公式与分类向量元素值的计算公式相同。

步骤4：关系融合即向量融合。将构建好的文本向量、分类向量和引用向量通过首尾相连的方式进行融合获得文献的多关系向量。假如给定某一篇文献的文本向量为$[\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3,\cdots,\boldsymbol{a}_n]$，分类向量为$[\boldsymbol{b}_1,\boldsymbol{b}_2,\boldsymbol{b}_3,\cdots,\boldsymbol{b}_n]$，引用向量为$[\boldsymbol{c}_1,\boldsymbol{c}_2,\boldsymbol{c}_3,\cdots,\boldsymbol{c}_n]$，则融合后的多关系融合向量为$[\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3,\cdots,\boldsymbol{a}_n,\boldsymbol{b}_1,\boldsymbol{b}_2,\boldsymbol{b}_3,\cdots,\boldsymbol{b}_n,\boldsymbol{c}_1,\boldsymbol{c}_2,\boldsymbol{c}_3,\cdots,\boldsymbol{c}_n]$。

第三层：聚类融合。首先计算第二步构建的多关系向量在空间上的距离和相似度，形成文本相似度矩阵，然后选择适合的如K−means、DIANA、OPTICS等聚类算法分别对论文、专利、基金项目进行聚类。聚类算法既可以选择相同的算法，也可以选择不同的算法。文本聚类后需要选择合适的聚类融合方法如余弦相似度、投票法等对3种类型数据形成的聚类簇进行聚类融合。

4 结语

多源信息融合在情报领域已得到一定的应用，相关研究多集中在数据类型融合和关系融合上。数据特征及类型融合主要侧重于信息融合过程中的多源异构信息的标准化，是多源信息融合的基础性工作。关系融合是多源信息融合的关键，在不同的应用场景中，各种关系类型的重要程度不同，承担的作用也不同。因此，面对不同的应用场景，如何选择合适的关系类型并进行融合是研究的重点内容之一。目前，在科技情报研究领域，对关系融合的研究较为丰富，在已有的关系融合方法中，简单的线性融合方法居多。从检索文献来看，在科技情报分析中，仅有文献[36]利用了聚类融合方法。对相关文献的梳理发现，目前科技情报研究中的多源信息融合深度不够，缺少统一的融合框架。本文认为，在科技情报研究中，利用多源数据进行情报分析是科技情报领域研究的关键问题之一。不同类型数据所包含的科技信息的侧重点各不相同。在科技情报研究中只有根据不同的研究目的而选取多种类型的数据进行信息融合，才能使科技情报分析的结果更加科学和客观。此外，随着计算机技术、机器学习等相关技术的不断发展，科技情报领域的多源信息融合方法和技术不再局限于数据类型本身的融合，更多强调的是融合的深度以及融合的模式。本文在梳理多源信息融合相关研究的基础上，总结现有研究方法的不足，以科技领域主题识别问题为例，从特征融合、关系融合、聚类融合3个层面提出并构建新的多源信息融合模式。该模式的提出为全面、深入地识别领域主题提供了借鉴与支持。

参考文献

[1] 化柏林.多源信息融合方法研究[J].情报理论与实践，2013, 36(11): 16−19.

[2] 李洋，赵鸣，徐梦瑶，等.多源信息融合技术研究综述[J].智能计算机与应用，2019, 9(5): 186−189.

[3] 靳杨，徐路路.基于本体语义增强和多源数据融合的石墨烯医学应用前沿探测[J].医学信息学杂志，2019, 40(2): 70−74, 85.

[4] 张维冲，王芳，赵洪.多源信息融合用于新兴技术发展趋势识别——以区块链为例[J].情报学报，2019, 38(11): 1166−1176.

[5] 化柏林.情报学三动论探析：序化论、转化论与融合论[J].情报理论与实践，2009, 32(11): 21−24, 41.

[6] 化柏林，李广建.大数据环境下多源信息融合的理论与应用探讨[J].图书情报工作，2015, 59(16): 5−10.

[7] ÁVILA R A, SENGOKU S.Tracing the knowledge−building dynamics in new stem cell technologies through techno−scientific networks[J].Scientometrics an international journal for all quantitative aspects of the science of science policy, 2017(112): 1691−1720.

[8] WANG H, ZHANG Z, WANG P. A situation analysis method for specific domain based on multi−source data

fusion[C]. Cham:Spring, 2018.
[9] 张付志, 刘华中.基于度量级融合的论文元数据提取方法[J].情报学报, 2013, 32(3): 235−243.
[10] 徐路路, 王芳.基于支持向量机和改进粒子群算法的科学前沿预测模型研究[J].情报科学, 2019, 37(8): 22−28.
[11] 冯佳, 穆晓敏, 王伟.面向研究前沿识别的载体−特征−关系融合模型研究[J].图书馆杂志, 2020, 39(9): 56−63.
[12] AMJAD T, DING Y, DAUD A, et al.Topic−based heterogeneous rank[J].Scientometrics, 2015, 104(1): 313−334.
[13] CALERO M C, NOYONS E. Combining mapping and citation network analysis for a better understanding of the scientific development: the case of the absorptive capacity field[J]. Journal of informetrics, 2008, 2(4): 272−279.
[14] 滕立.基于超网络的作者−机构−国家混合共现网络研究[J].情报学报, 2015, 34(1): 9.
[15] BEI W, HORLINGS E, ZOUWEN M, et al.Mapping science through bibliometric triangulation: an experimental approach applied to water research[J].Journal of the association for information science & technology, 2017(68): 724−738.
[16] ZHANG Y, SHANG L, HUANG L, et al.A hybrid similarity measure method for patent portfolio analysis[J]. Journal of informetrics, 2016, 10(4): 1108−1130.
[17] 谭晓, 李辉.基于多源数据知识融合方法的研究前沿识别[J].现代情报, 2019, 39(8): 29−36.
[18] TANG J, JIN R, ZHANG J.A topic modeling approach and its integration into the random walk framework for academic search[C]//Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008). Washington, DC: IEEE Computer Society. 2008.
[19] TANG J, JIN R, ZHANG J, et al.Topic level expertise search over heterogeneous networks[J].Machine learning, 2011, 82(2): 211−237.
[20] XU S, SHI Q, QIAO X, et al.Author−topic over time (AToT): a dynamic uers' interest model[J].Lecture notes in electrical engineering, 2014, 274: 239–245.
[21] XU S, SHI Q, QIAO X, et al.A dynamic users' interest discovery model with distributed inference algorithm[J]. International journal of distributed sensor networks, 2014 (1): 1−11.
[22] 史庆伟, 乔晓东, 徐硕, 等.作者主题演化模型及其在研究兴趣演化分析中的应用[J].情报学报, 2013, 32(9): 912−919.
[23] DU Y P, CQ Y, LI N.Using heterogeneous patent network features to rank and discover influential inventors[J].信息与电子工程前沿(英文版), 2015, 16(7): 568−578.
[24] XU H Y, YUE Z H, WANG C, et al.Multi−source data fusion study in scientometrics[J].Scientometrics, 2017, 111(2): 773−792.
[25] STREHL A, GHOSH J. Cluster ensembles: a knowledge reuse framework for combining multiple partitions [J]. Journal of machine learning research, 2002, 3(3): 583−617.
[26] 张洪. 聚类集成算法在客户细分中的研究及应用[D].合肥: 安徽大学, 2016.
[27] TUMER K, AGOGINO A K.Ensemble clustering with voting active clusters[J].Pattern recognition letters, 2008, 29(14): 1947−1953.
[28] 张静静, 杨燕, 王红军, 等.一种新的软聚类投票法及其并行化实现[J].中国科学技术大学学报, 2016, 46(3): 173−179.
[29] 陈晓云, 陈刚.基于最大内聚度基准的加权投票聚类集成[J].控制与决策, 2014, 29(2): 236−240.
[30] 李锋, 李寿梅.基于证据理论的聚类集成方法[J].南京信息工程大学学报(自然科学版), 2019, 11(3): 332−339.
[31] HUANG D, LAI J H, WANG C D.Combining multiple clusterings via crowd agreement estimation and multi−granularity link analysis[J].Neurocomputing, 2015, 170 (25): 240−250.
[32] LAI J H, WANG C D, DONG H, et al.Ensembling over−segmentations: from weak evidence to strong segmentation[J]. Neurocomputing, 2016, 207: 416−427.
[33] 毕凯, 王晓丹, 邢雅琼.基于模糊测度和证据理论的模糊聚类集成方法[J].控制与决策, 2015, 30(5): 823−830.
[34] 王田.基于超图划分的高维数据聚类方法研究[D].兰州: 兰州大学, 2018.
[35] WANG X, YANG C Y, ZHOU J.Clustering aggregation by probability accumulation[J].Pattern recognition, 2009, 42(5): 668−675.
[36] 许晓阳, 郑彦宁, 刘志辉.论文和专利相结合的研究前沿识别方法研究[J].图书情报工作, 2016, 60(24): 97−106.
[37] 邵长龙, 孙统风, 丁世飞.基于信息熵加权的聚类集成算法[J].南京大学学报(自然科学), 2021, 57(2): 189−196.