



开放科学  
(资源服务)  
标识码  
(OSID)

# 基于多源信息融合的专家画像方法研究

李贞亮<sup>1</sup> 吴振峰<sup>1</sup> 高嵩<sup>1</sup> 张宇飞<sup>1</sup> 刘晗<sup>2</sup> 李龙豪<sup>2</sup>

1. 中国科学技术信息研究所 北京 100038;
2. 天津理工大学计算机科学与工程学院 天津 300384

**摘要:** [目的/意义] 在项目评审、人才评价、机构评估改革等科研评价工作中,同行评议专家的遴选是最重要的环节,专家遴选的准确性在很大程度上影响评议结果的质量,而这种准确性高度依赖专家标签的质量,专家画像研究是科研评议面临的重要问题之一。[方法/过程] 提出一种基于多源信息融合的专家画像构建方法。该方法通过融合多源异构数据提取不同角度的专家科研指标,使用 BERT 预训练模型和 Bi-LSTM 进行实体抽取,获得更加全面的专家标签特征。[局限] 需要一定量的数据用于提前预训练 BERT 模型。[结果/结论] 该方法能够为专家刻画更加丰富的画像,为专家画像构建方法研究工作提供参考和借鉴。

**关键词:** 专家画像; 专家遴选; 多源信息; 数据融合

**中图分类号:** G35; TP391

## Research on Method of Expert Portrait Based on Multi-source Information Fusion

LI Zhenliang<sup>1</sup> WU Zhenfeng<sup>1</sup> GAO Song<sup>1</sup> Zhang Yufei<sup>1</sup> LIU Han<sup>2</sup> LI Longhao<sup>2</sup>

1. Institute of Scientific and Technical Information of China, Beijing 100038, China;
2. School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China

**Abstract:** [Objective/Significance] The selection of peer review experts is the most important part in scientific research evaluation, such as project review, talent evaluation, and institutional evaluation reform. The accuracy of expert selection greatly affects the quality of evaluation results, and this accuracy highly depends on the quality of expert labels. Therefore, expert portrait research is one of the important issues facing scientific research evaluation. [Methods/Processes] We propose a method for constructing expert portraits based on multi-source information fusion. This method extracts expert research indicators from different perspectives by integrating multi-source heterogeneous data, and uses BERT models and Bi-LSTM for entity extraction

**基金项目** 中国科学技术信息研究所创新研究基金青年项目“基于多源数据模糊融合的专家标签树构建方法研究”(QN2024-16)。

**作者简介** 李贞亮(1995-), 硕士, 助理研究员, 主要研究方向为软件工程、科技情报分析、科技政策研究; 吴振峰(1992-), 通信作者, 博士, 副研究员, 主要研究方向为自然语言处理与智能情报, E-mail: wuzf@istic.ac.cn; 高嵩(1998-), 硕士, 研究实习员, 主要研究方向为软件工程、科技情报分析; 张宇飞(1997-), 硕士, 研究实习员, 主要研究方向为软件工程、科技情报分析; 刘晗(2000-), 硕士研究生, 主要研究方向为深度学习、图像处理; 李龙豪(2000-), 硕士研究生, 主要研究方向为深度学习、图像处理。

**引用格式** 李贞亮, 吴振峰, 高嵩, 等. 基于多源信息融合的专家画像方法研究[J]. 情报工程, 2025, 11(5): 48-57.

to obtain more comprehensive expert label features. [Limitations] This method requires pre-training the BERT model in advance and requires a certain amount of data. [Results/Conclusions] This method could depict more diverse portraits providing reference and inspiration for research on expert portrait construction methods.

**Keywords:** Expert Portrait; Expert Selection; Multi-source Information; Data Fusion

## 引言

科技评价是对科技活动产生的效果和影响的认识过程，在推动国家科技事业持续健康发展、优化科技资源配置、提高科技管理水平、激励和引导科技创新等方面具有重要作用。

同行评议制度被广泛应用于项目评审、人才评价、机构评估改革等科研评价工作中，专家遴选是这项制度的一种基本方法，其准确性很大程度上决定了评议结果的质量，而这种准确性高度依赖专家标签的质量。“小同行”无论在研究领域还是科研经历中都与评议对象具有更高的契合度。但是随着学科分级的不断细化、研究方向不断增多，交叉研究、跨学科研究成为趋势，遴选合适的评审专家开展同行评议变得更加困难，如何对科研专家开展精准画像成为科研管理工作的重要问题。现有专家库、专家画像构建研究很注重专家标签的刻画，但是广泛存在专家标签描述颗粒度较粗，标签更新不及时、不准确、不全面等问题。科学的专家标签能够准确地描述专家的研究情况，同时有助于同行评议活动的开展，解决评议过程中“同行遴选难”的问题。因此，专家标签构建方法研究无论是在理论层面还是实践层面都具有重要意义。

本文提出一种基于多源信息融合的专家画像构建方法。该方法通过对论文、专利、科技

报告等不同数据信息进行特征类型融合和数据加权，提取不同角度的科研指标，然后使用BERT（Bidirectional Encoder Representation from Transformers）预训练模型进一步提取特征信息，通过Bi-LSTM（Bi-directional Long Short Term Memory）和CRF（Conditional Random Field）模型进行实体抽取，获得更加全面的专家标签特征用于专家画像构建。该方法能够充分发挥多源数据互相补充的作用，提高构建标签的真实性、可靠性与正确程度，减少信息的模糊性、不确定性和语义偏差。

## 1 相关研究概述

### 1.1 专家画像研究现状

专家标签是描述专家研究领域和研究方法的一组关键词组，专家画像就是把各类信息进行抽象提炼后所形成的特征标签组合。专家标签是专家画像的重要组成部分，其作用在于刻画专家研究属性，为遴选评议专家等工作提供判断依据。国内外学者对标签构建研究颇多，并将标签构建应用于画像研究中。孟巍等<sup>[1]</sup>指出，科研人员画像就是把各类信息进行抽象提炼后所形成的特征标签组合。马超<sup>[2]</sup>在此基础上提出基于隐含狄利克雷分布（Latent Dirichlet Allocation, LDA）的半监督学习用户画像模型。王锐杰<sup>[3]</sup>则采用条件随机场模型来抽

取科学学者的相关特征,进而实现对科学学者的画像构建。袁莎等<sup>[4]</sup>提出运用概率因子图模型识别学者间关系,同时借助主题模型对学者特征进行聚类,实现对学者信息的挖掘与分析。而在国外,Blei等<sup>[5]</sup>等提出了一个围绕文档、主题以及词构建的LDA模型,以此来提高模型的性能表现与鲁棒性。除此之外,Tang等<sup>[6]</sup>还提出了一种专门针对Web网页的标签组合算法。

总的来说,目前关于画像的研究有很多,这些研究通常将每个标签都视为独立的、不可再分的原子标签。这种标签体系颗粒度较粗、不成体系,很难刻画不断细化的研究方向,满足“小同行”精准化专家遴选要求。

## 1.2 多源信息融合研究现状

多源数据融合源于军事领域研究,现如今已广泛应用于情报分析、地理空间系统等领域。在情报领域,随着互联网和信息技术的快速发展,传统单一数据研究方法很难挖掘整合深层有效信息,无法有效满足科技决策部门对全方位、专业化科技情报的实际需求。多源数据融合作为一项综合性技术手段和分析方法,能够实现不同源头的采集数据有机整合、相互共享,相较于依靠单一数据源而言,具备更高的可靠性、科学性和有效性,能够挖掘出更为全面且深入的信息。

在当今的学术研究中,多源信息融合工作吸引了众多学者的目光,相关方法和技术探索取得诸多颇具价值的成果。Khaleghia等<sup>[7]</sup>专注于数据融合领域,对诸多方法进行梳理,清晰阐释了相关的基础概念,为该领域的后续发展

铺就了理论基石。Chopra等<sup>[8]</sup>独辟蹊径,创新性地提出运用特制的软件工具和算法模型来拆解多源信息融合难题,为精准决策提供可靠依据。陈金海<sup>[9]</sup>在科研成果信息融合特性的深度研究基础上,总结提炼情报数据融合技术的科学处理流程,为相关行业情报处理工作提供实操指南。李广建等<sup>[10]</sup>深入分析大数据时代情报研究,明确指出当下的情报研究范畴应横跨各个领域,打破单一领域的局限。郑彦宁等<sup>[11]</sup>则精准定位“多源信息”与“多元方法”协同优势,切实提升情报研究的应用价值。传统的多源信息融合方法只聚焦于刻画标签的真实性,很少考虑标签的多样性、实时性。多源信息融合分析在专家画像构建应用更具优势。如何促进多源信息融合,充分利用多来源数据信息,是一个值得深入研究的方向。

## 2 基于多源信息融合的专家画像

本文提出基于多源数据信息融合的方式构建专家画像,将用户画像理论与科技专家应用服务现实需求相结合,充分利用大模型等技术工具,通过动态融合互联网多源数据特征信息综合刻画科技专家画像,从指标体系设计构建、多源数据采集处理、多源特征信息融合、专家画像模型构建以及专家应用与服务等方面对科技专家画像模型展开研究。具体阶段和构建流程如图1所示。

### 2.1 专家画像指标体系构建

指标体系建设是专家画像建设的首要任务,本文围绕专家基本信息、科研信息和其他信息等三个维度构建专家画像指标体系。其中,个

人信息维度侧重于描述专家的基本特征信息(如性别、年龄、单位等), 科研信息、其他信息

维度侧重展示不同信息源、不同类型的专家数据信息。

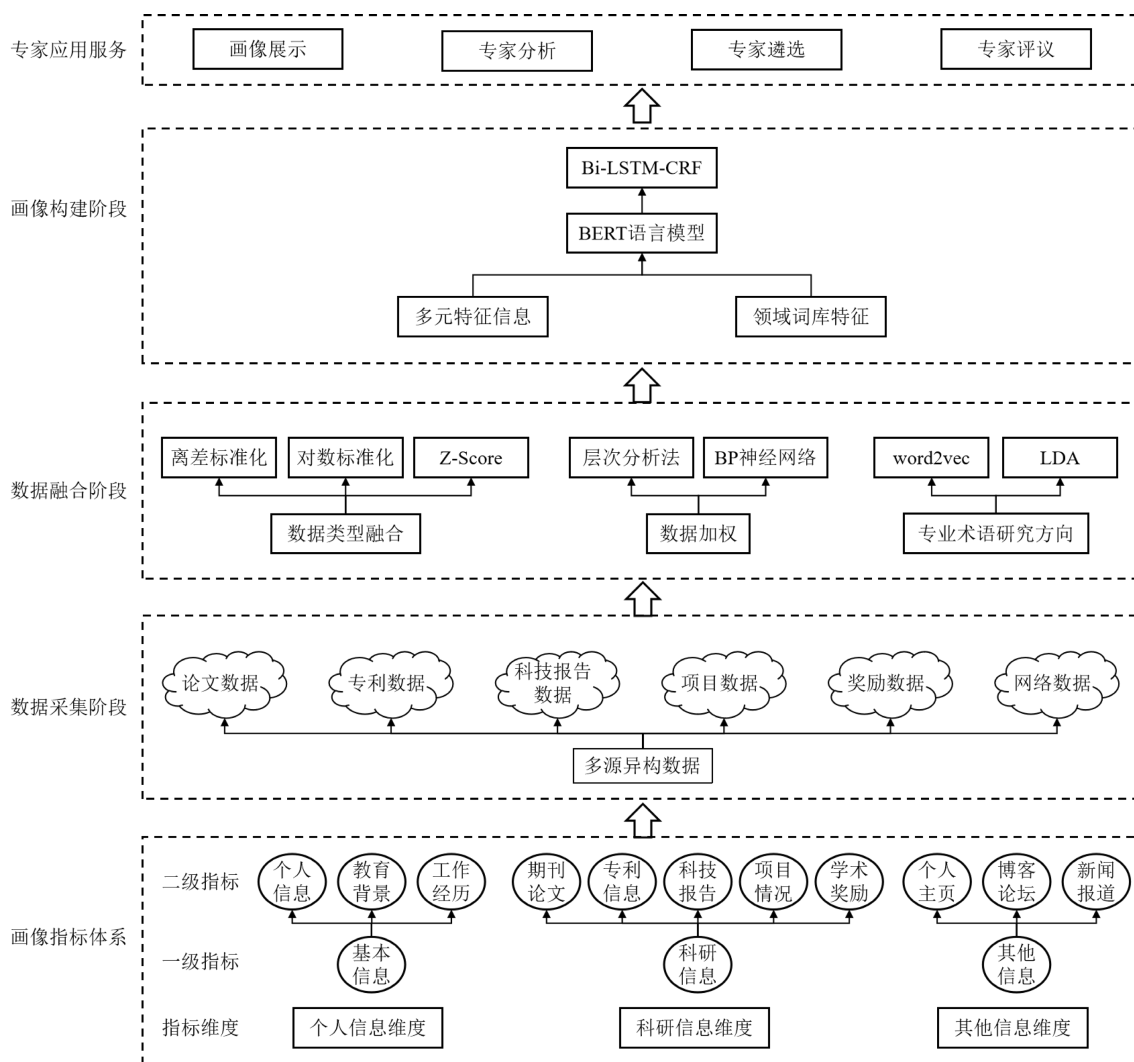


图1 基于多源信息融合的专家画像构建模式

## 2.2 多源数据采集

专家画像指标体系对于精准了解专家群体特征、助力多领域决策有着至关重要的作用。为构建完整且精准的专家画像，本文根据设置的专家画像指标体系，从多渠道获取较为全面的数据，一方面通过官方渠道获取论文、专利等结构化数据信息，另一方面通过分布式爬虫算法采集非结构化数据。本文结合多源异构数

据开展实证分析。成功获取各维度的指标数据后，需要在开展多源数据融合工作前对其进行数据清洗工作。本文所涉及的数据清洗流程包括三个环节：异常数据检测、缺失值排查和处理、异常值的筛查和处理。

## 2.3 多源信息融合

本文扩展数据采集范围，使用多渠道异构数据刻画专家画像，提升专家分析的准确性和

全面性。不同指标数据的来源各异，不同成果的存储形式和数据类型存在显著差异，这使得多源异构数据在开展融合前需要进行必要的预处理。本文重点聚焦于数据融合转换和数据加权这两个问题展开。

(1) 数据类型融合。首先根据指标体系将采集的多源数据信息转化为定量指标信息。由于不同来源的定量指标彼此存在量纲差异，需要通过离差标准化、对数标准化和 Z-SCORE 算法等，实现定量数据的无量纲化。无量纲化后，不同单位或量级的数据指标可以进行相互比较，便于后续加权分析和处理。其次，将采集的低密度数据转化为定性指标信息，去除无实际意义的噪声数据。博客、论坛、新闻报道等公开信息以定性指标分析为主。

在定性一定量指标数据的转换处理方面，可以使用训练好的 word2vec 模型对自然语句进行词向量化。对于由  $n$  个单词组成的句子  $S=\{c_1, c_2, \dots, c_n\}$ ，每个单词对应的词向量可以表示为：

$$e_i^c = \text{word2vec}(c_i) \quad (1)$$

其中， $c_i$  表示句子  $S$  中的第  $i$  个单词， $e_i^c$  表示单词  $c_i$  对应的词表征向量。

(2) 数据加权问题。根据上述方法，可以从不同来源的数据文本中提取出多组定性和定量的指标描述。一个专家往往只需要一组指标体系进行最终的刻画展示，这就涉及数据指标加权问题。在完成不同类型的指标数据融合之前，需要确定各个指标之间的权重信息、多源数据信息的重要性。

通过多源数据形成一个统一的特征画像，首先考虑的是不同类型资源的赋权分析问题。

论文发表、研究项目申请往往具有严格审核流程，在刻画专家的研究属性方面权威性最高；个人主页通常是专家本人维护，具有很强的准确性和时效性。其次，数据资源的时间属性是判断数据价值的重要因素，尤其是揭示专家特征的资源存在较大噪声的情况下，近期的资源对专家画像特征的揭示要早于较早年份。此外，词频也是判断一个语句中不同单词重要程度的关键指标。

对于由  $m$  个句子构成的一组多源异构语句  $\{S_1, S_2, \dots, S_m\}$ ，对于其中第  $i$  个单词  $w$ ，要想获取其对应的多源特征信息，需要综合词频、数据年份和数据来源等 3 方面因素考虑，然后得到每个单词的加权表征向量，即：

$$e_i^w = \begin{cases} \sum_{j=1}^m f_j(w) t_j r_j \cdot e^c, & \{S_1, S_2, \dots, S_m\} \neq \emptyset \\ e_{default}, & \{S_1, S_2, \dots, S_m\} = \emptyset \end{cases} \quad (2)$$

其中， $e_i^w$  表示融合不同来源语句  $\{S_1, S_2, \dots, S_m\}$  的加权词表征向量， $e^c$  表示单词  $w$  通过 word2vec 模型向量化的原始表征向量， $f_j(w)$  表示单词  $w$  在语句  $S_j$  中的词频， $t_j$  表示考虑数据年份的权重指标，根据语句  $S_j$  的年份给出对应的权重值， $r_j$  表示考虑数据资源类型的权重指标，根据语句  $S_j$  的语料权威性给出对应的权重值， $e_{default}$  表示不存在多源异构数据的情况下的词向量，可以使用零向量表示，也可以是所有单词词向量的均值。通过对专家各个维度指标的加权计算进行多源数据信息融合，为下一步提取专家特征信息、构建专家画像做好数据准备。

(3) 领域词库提取。为了提取更多的特征信息用于专家画像构建，本文还提出使用 LDA 主题抽取模型等直接对多源异构数据进行专业术语、专家领域词汇提取，并通过构建相关领

域专业词库辅助分词和处理词嵌入。最后将加权词表征向量  $e_i^w$  和原始词表征向量  $e_i^c$ ，通过向量拼接进一步实现多源特征信息融合  $E_i$ ，尽可能保留更多的专家特征信息，通过多个数据来源综合利用和深度学习算法模型完成专家画像构建，即：

$$E_i = e_i^c \oplus e_i^w \quad (3)$$

## 2.4 专家画像构建

实体抽取 (Entity Extraction) 是自然语言处理领域的一项基准任务，主要是从文本数据中自动识别具有特定语义的实体名称，例如姓名、地点等信息。在专家画像构建时，实体抽取是

一项重要的技术手段，主要用于各类专业术语、主题词和关键词抽取。为了更加全面、准确地刻画专家画像，本文使用通过数据挖掘产生的大量特征信息，作为实体抽取模型的基本语料。

考虑到不同算法在处理文本信息时各有优势，使用多种机器学习方法组合的方式能够克服单一算法模型的局限性<sup>[12]</sup>。如图2所示，本文提出使用经过预训练的BERT模型挖掘词语间的关联关系，通过注意力机制和非线性映射捕获不同来源数据的长距离依赖信息。在此基础上使用经典的循环神经网络模型 (Recurrent Neural Network, RNN) 和 CRF 学习时序序列数据特征。

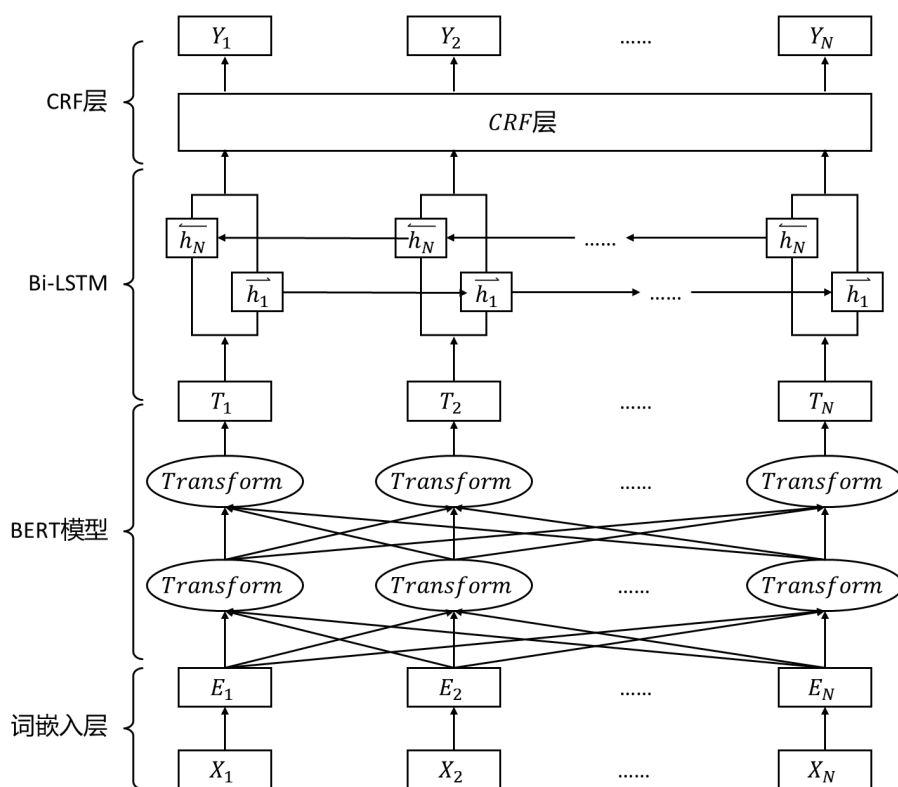


图2 BERT模型和Bi-LSTM-CRF算法模型

为了更精确地建立全方位、多维度的专家画像，首先使用上述多源信息融合的方法将赋予权重信息的特征数据  $S=\{c_1, c_2, \dots, c_N\}$  进行词

嵌入和特征向量拼接，组合多源特征信息，得到初始向量序列  $E=\{E_1, E_2, \dots, E_N\}$ ， $N$  为特征向量的长度，即 BERT 模型的输入长度。

BERT 模型是基于 Transformer 的双向编码器表征模型，模型在处理某一个词时，可以综合利用句子前后字词关系信息，获得这个词的语义信息<sup>[13-14]</sup>。由于文本语料信息较少，为防止神经网络出现过拟合现象，本文使用 12 层的 Transformer Encoder 搭建 BERT 模型网络。将初始向量序列  $E$  送入预训练好的 BERT 模型后，便可以获得每个输入向量对应的特征信息  $T$ 。

$$T_i^E = BERT(E_i) \quad (4)$$

其中， $E_i$  表示第几个字符， $T_i^E$  表示特征向量  $E$  第  $i$  个字符的训练结果。

循环神经网络是一种流行的文本处理模型。相较于传统文本处理方法中文本特征稀疏和语义敏感的问题，循环神经网络能够更为有效处理序列数据，发现文本数据的隐藏特征信息。Bi-LSTM 是一种基于长短期记忆网络 (LSTM) 的时间序列预测方法，如图 2 所示。它结合了双向模型和 LSTM 的门控机制，由 2 个独立的 LSTM 网络构成。当 Bi-LSTM 处理序列数据时，输入序列会分别以正序和逆序输入到 Bi-LSTM 网络中进行特征提取，并将输出向量（即提取后的特征向量）拼接后形成的输出向量作为该时间的最终输出。

Bi-LSTM 的正向输出  $\vec{h}_t$  与反向输出  $\overleftarrow{h}_t$  的定义如下，最终输出结果  $\bar{h}_t$  定义如式 (7) 所示：

$$\vec{h}_t = LSTM(T_i, \vec{h}_{t-1}), t = 1, 2, \dots, N \quad (5)$$

$$\overleftarrow{h}_t = LSTM(T_i, \overleftarrow{h}_{t-1}), t = 1, 2, \dots, N \quad (6)$$

$$\bar{h}_t = w\vec{h}_t + v\overleftarrow{h}_t + b_t, t = 1, 2, \dots, N \quad (7)$$

其中： $w$ 、 $v$ 、 $b$  分别为正向输出权重矩阵、

反向输出权重矩阵和偏置。

最后将获得的嵌入特征送入条件随机场 CRF 网络层中，即可获得实体抽取的结果。

$$Y_i = CRF(\bar{h}_i), i = 1, 2, \dots, N \quad (8)$$

## 2.5 专家应用与服务

根据上文所述的多源信息融合方法和画像构建方法，生成专家画像所需的各项指标信息。将构建的画像信息可视化地展现给用户，辅之以其他技术手段实现专家画像的应用价值，包括：

(1) 画像展示。通过多源信息融合方法，对特定领域专家进行多维度、综合性描述，形成立体、动态的专家模型。通过可视化方式直观展示，以便更好地理解、识别、管理和利用专家资源。

(2) 专家分析。建立专家评价指标体系，基于融合多源信息特征的画像开展专家分析，了解行业领域的专家人才结构、产业分布、发展趋势等，为行业的人才培养、政策制定等提供信息支持。

(3) 专家遴选。在组织专家使用或专家推荐工作中，可以将专家遴选问题转化为关键词匹配问题，依据专家画像进行精准匹配，确保推荐的专家能够满足推荐使用实际，提供有效的咨询方案。

(4) 专家评议。根据专家画像和遴选推荐，将指定领域经验丰富、专业知识背景深厚的专家用于科研学术、项目评审、人才评价等领域的评估评议工作。依靠与评议对象相关的领域专家，按照评议标准和程序开展专业客观的评价和审议。

### 3 实证分析

#### 3.1 数据来源

为了突出本文提出方法的可行性和效果，在数据选择上，拟选择交叉学科领域、跨领域研究专家进行研究，同时考虑到不同领域成果信息、数据量不同，拟优先选择热点领域研究开展分析。本文从 Web of Science 等大型数据库和学术期刊网站等渠道获取一批新能源汽车领域的专家数据，剔除重复文本和无效数据，构建专家多源异构数据集，形成包括 203 名专家的 21850 篇文档。这些文件按照 6:3:1 的比例

划分为训练集、验证集和测试集。此外，本文使用的 BERT 网络是一个深度学习预训练模型。本文通过其他领域的论文和专利文件库开展模型预训练和参数微调。

#### 3.2 模型训练

在进行多源信息融合时，不同年份、不同来源的数据资源需要赋予不同的权重值。本文经过多轮调试，按照表 1 设置数据资源年份  $t_j$  和类型  $r_j$  的权重信息，发布时间更近、信息来源更权威的资源都赋予更高的权重值，通过加权公式实现不同类型的指标数据融合。

表 1 不同年份数据资源权重设置标准

数据资源年份	3年内发布的资源	3~5年前发布的资源	5~10年前发布的资源	10~20年前发布的资源	20~30年前发布的资源	发布超30年的资源		
权重	5	4	3	1	0.5	0		
数据资源类型	期刊论文	专利信息	科技报告	项目情况	学术奖励	个人主页	博客论坛	新闻报道
权重	5	4	3	3	4	3	1	1

#### 3.3 结果分析

为了验证算法模型的抽取性能，本文设置了一系列对比实验，将本文所提方法和基于机器学习的标签构建算法 LDA 模型，K 近邻 (K-Nearest Neighbor, KNN)，基于深度学习的标签构建算法卷积神经网络 (Convolutional Neural Network, CNN)、BiLSTM+Attention 模型在专家数据集上分别进行了实验，本文使用通用评价指标 (准确率、精准率、召回率、F1 Score) 评价模型效果，对比实验结果如表 2 所示。相较于其他模型，本文所提基于多源信息融合的标签构建方法可以充分考虑不同来源数据的互相补充作用，更精准地构建专家标签，

其 F1 Score 值为 93.15%，高于其他模型。

表 2 对比实验结果

模型	准确率	精准率	召回率	F1 Score
LDA	79.51%	78.70%	79.14%	79.37%
KNN	80.20%	81.10%	82.25%	81.56%
CNN	87.25%	88.53%	88.44%	87.37%
BiLSTM+Attention	90.68%	90.12%	89.78%	89.68%
Ours	92.11%	92.35%	92.42%	93.15%

#### 3.4 专家画像

构建专家画像就是将专家特征标签化。本文使用语义化的短文本标签对专家群体的个人

属性、科研属性和其他属性进行描述，全面展示专家的特征信息，为遴选评议专家等工作提供判断依据。

本文从新能源汽车领域的 203 位专家中随机抽取两位专家进行基本特征画像，专家姓名用 ID 编号代替，从多个维度展示专家的特征信

息。多源数据信息能够从多个视角反映专家的研究内容和方向。如图 3、图 4 所示，专家画像中为每位专家刻画了一个包括学习经历、工作地点和研究方向的侧画像。画像显示新能源汽车领域专家研究内容较为广泛，涵盖电气工程、控制工程、机械等领域。



图 3 专家信息基本界面

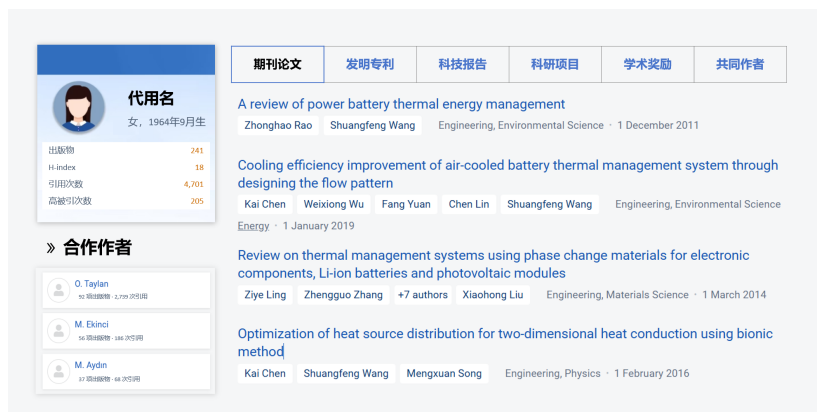


图 4 专家多源异构数据成果展示界面

## 4 结语

专家画像刻画的核心在于专家标签构建，完善高效的专家标签体系有助于同行评议活动的开展，解决评议过程中“同行遴选难”的问题。本文结合深度学习算法提出一种基于融合多源异构数据信息的专家画像构建方法。这种方法能够提高构建标签的真实性、可靠性与正确程

度，减少信息的模糊性、不确定性和语义偏差，同时可以有效利用互联网上广泛存储的大数据信息，充分发挥多源数据互相补充的作用，减少信息资源浪费。专家标签构建方法研究无论是在理论层面还是实践层面都具有重要意义。此外，现有专家画像构建研究广泛存在专家标签描述颗粒度较粗，标准不统一的问题，这些问题将进一步纳入后续研究考虑。

## 参 考 文 献

- [1] 孟巍, 吴雪霞, 李静, 等. 基于大数据技术的电力用户画像 [J]. 电信科学, 2017, 33(S1): 15-20.
- [2] 马超. 基于主题模型的社交网络用户画像分析方法 [D]. 合肥: 中国科学技术大学, 2017.
- [3] 王锐杰. 基于多源信息融合的科研学者画像及应用研究 [D]. 成都: 电子科技大学, 2020.
- [4] 袁莎, 唐杰, 顾晓韬. 开放互联网中的学者画像技术综述 [J]. 计算机研究与发展, 2018, 55(9): 1903-1919.
- [5] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of machine Learning research, 2003(3): 993-1022.
- [6] TANG J, YAO L, ZHANG D, et al. A combination approach to web user profiling[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2010, 5(1): 1-44.
- [7] KHALEGHI B, KHAMIS A, KARRAY F O, et al. Multisensor data fusion: a review of the state-of-the-art[J]. Information fusion, 2013, 14(1): 28-44.
- [8] CHOPRA K, HAIMSON C. Information fusion for intelligence analysis[C]// Proceedings of the 38th Annual Hawaii International Conference on System Sciences. IEEE, 2005: 111a-111a.
- [9] 陈金海. 关于情报信息融合处理方法的研究 [J]. 情报杂志, 2003(3): 63-64.
- [10] 李广建, 杨林. 大数据视角下的情报研究与情报研究技术 [J]. 图书与情报, 2012(6): 1-8.
- [11] 郑彦宁, 刘志辉, 赵筱媛, 等. 基于多源信息与多元方法的产业竞争情报分析范式 [J]. 情报学报, 2013, 32(3): 228-234.
- [12] NASAR Z, JAFFRY S W, MALIK M K. Textual keyword extraction and summarization: state-of-the-art[J]. Information Processing and Management, 2019, 56(6): 102088.
- [13] 曾子明, 陈思语. 基于 LDA 与 BERT-BiLSTM-Attention 模型的突发公共卫生事件网络舆情演化分析 [J]. 情报理论与实践, 2023, 44(4): 1-13.
- [14] 刘逸伦, 黄微, 张晓君, 等. AIGC 赋能的科技情报智能服务: 特征、场景与框架 [J]. 现代情报, 2023, 43(12): 88-99.