大数据管理与应用专业岗位能力需求分析及人才培养优化策略*
——基于招聘信息的文本挖掘

孙倬[1,2,3]　李嘉欣[1]　鲁晓明[1,2]　杨瑞仙[1,2,3]
（1.郑州大学信息管理学院，郑州 450001；2.河南省数据治理研究中心，郑州 450001；
3.河南省“双一流”建设研究中心，郑州 450001）

摘要：在数字化转型与产教融合不断深化的背景下，大数据管理与应用专业存在人才供给与岗位需求错配问题。本研究选取前程无忧网站在线招聘文本，统计分析学历、经验、职位类型与地区分布等岗位基本需求，并采用潜在狄利克雷分配（Latent Dirichlet Allocation，LDA）模型识别能力主题，结合技能抽取与能力共现刻画岗位能力需求。研究发现，岗位基本需求呈现四个方面特征：学历要求以本科为主；近半数岗位面向的是有工作经验者；岗位地区分布集中在核心城市与区域中心城市；岗位类型可归纳为数据类、工程研发类、业务管理类。岗位能力需求主要围绕智能化项目管理与协作、数据驱动的业务优化分析、客户关系与销售运营管理、数据系统开发与数据库技术运用四个能力维度展开，且不同类型的岗位对人才能力的需求存在显著差异。基于需求侧视角对大数据管理与应用专业岗位能力需求的分析，本研究提出针对性人才培养优化策略，为专业人才培养提供实践依据，助力国家大数据产业发展。

关键词：大数据管理与应用；招聘信息；文本挖掘；能力组合；需求分析；人才培养
中图分类号：G642　DOI：10.3772/j.issn.1673-2286.2026.06.004
引文格式：孙倬，李嘉欣，鲁晓明，等. 大数据管理与应用专业岗位能力需求分析及人才培养优化策略：基于招聘信息的文本挖掘[J]. 数字图书馆论坛，2026，22（6）：39-48.

在数智化转型加速的背景下，大数据已被看作是战略性基础资源和组织获得竞争优势的关键因素[1]。随着大数据产业的快速兴起，企业对数据分析、数据治理与数据驱动决策人才的需求日益凸显，但现有人才供给难以满足复合型岗位对人才能力的要求。在此背景下，高校在培养社会急需的大数据人才方面发挥着重要作用[2]。2017年，教育部新增备案本科专业“大数据管理与应用”（代码120108T），在西安交通大学、哈尔滨工业大学、东北财经大学、南京财经大学、贵州财经大学这5所高校设立该本科专业，修业年限4年，学位授予门类为管理学[3]。自首批5所高校开设大数据管理与应用专业以来，该专业布点快速扩展，截至2022年12月，该专业的培养高校数量达到183所[4]；截至2024年10月，除港澳台地区外，全国共有252所高校开设大数据管理

收稿日期：2026-04-21
*本研究得到2025年度郑州大学研究生教育研究项目“生成式人工智能在研究生论文写作指导与学术规范教育中的应用研究”（编号：YJSJU2025075）、2025年度郑州大学研究生优质课程项目“用户需求分析与系统设计”（编号：SFKC202538）、2025年度郑州大学课程思政教育教学示范课项目“管理学原理”（编号：2025ZZUKCSZ061）资助。

与应用本科专业[5]。

在政策方面，2015年，国务院发布《促进大数据发展行动纲要》[6]，推动大数据上升为国家战略；2021年，工业和信息化部发布的《“十四五”大数据产业发展规划》指出“数据是新时代重要的生产要素，是国家基础性战略资源”“到2025年，大数据产业测算规模突破3万亿元”“加快人才培养”[7]，为大数据专业人才培养提出了新的需求。在人才需求方面，《中华人民共和国职业分类大典（2022年版）》[8]中首次标注97个数字职业，显示数字化人才队伍正处于快速增长阶段，其中包含大数据工程技术人员，表明该职业已得到国家标准化认定，具备独立、稳定的工作内容与从业群体，并被纳入全国职业统计体系，可为职业培训、人才规划、产业政策制定提供依据。然而，大数据管理与应用专业建设存在针对性不强、培养质量不高、专业特色不明显等问题[2]，且大多数高校的课程体系和培养目标仍存在不完全匹配的问题[4]。因此，如何精准把握岗位需求，推动人才培养与岗位能力需求的有效匹配，已成为大数据管理与应用专业人才培养的重要问题。

现有研究[9–10]表明，招聘信息能较为直接地反映用人单位对知识、技能与素质的组合需求与期望，是开展人才需求分析的重要数据来源。但针对大数据管理与应用专业的系统性岗位能力需求研究仍较少，且部分研究停留于描述性统计或粗粒度主题归纳，而人才能力需求分析与人才培养之间缺乏清晰的映射关系。能力是个体完成岗位任务所需要的知识、技能及相关职业素养的综合体现。其中，技能是能力在具体岗位要求中的重要表现形式，也是招聘文本中能够直接识别和抽取的可观测要素。为此，本研究从招聘信息入手，探讨大数据管理与应用相关岗位的能力主题维度及其结构特征，分析不同能力组合的差异，将需求侧的能力组合转化为供给侧可落地实施的课程群、项目任务与学习成果评价，旨在建立岗位能力需求与专业人才培养之间的映射关系，为该专业课程体系优化、实践教学设计等提供依据。

1　相关研究

1.1　基于在线招聘文本的人才需求分析研究

在线招聘文本包含岗位类别、学历、工作经验、岗位职责及任职要求等信息，是分析人才需求的重要数据来源。现有研究主要采用内容分析法和文本挖掘法两类方法。①内容分析法通过分词技术和词频统计等方式识别人才需求的基本特征。王伟[11]基于对招聘广告的文本分析，从招聘企业、职位、任职要求三个维度对金融科技人才需求进行统计和分析，并通过相关度计算来揭示岗位与任职要求的内在关系。贾东琴等[12]运用文本分析方法对我国公共图书馆招聘信息进行分析，从岗位类别、学历背景与能力要求等维度解析其人才需求特征。胡忠义等[9]以招聘信息为数据来源，对商务智能岗位的基本情况、核心技能和员工素质进行分析，并采用相关性分析来探究工作经验和学历对薪资待遇的影响。②文本挖掘法通过实体抽取、术语抽取、主题建模等技术，从语义层面识别岗位所需的知识、技能及其关联关系[13]。袁毅等[14]通过实体抽取并构建关系网络分析人工智能领域人才需求。王东波等[15]基于招聘公告构建深度学习模型，自动抽取数据科学招聘实体。俞琰等[16]利用依存句法和术语抽取方法识别岗位技能信息，以明确当前就业市场对人才的技能需求。主题建模属于文本挖掘的一种具体技术路径，潜在狄利克雷分配（Latent Dirichlet Allocation，LDA）、LDA2Vec和BERTopic等模型可用于识别招聘文本中的潜在需求主题[17]。岳铁骐等[18]采用LDA模型识别岗位需求主题，并结合Word2Vec与依存句法分析互联网行业人才能力需求及其关联关系。总体来看，基于在线招聘文本的人才需求分析已由岗位基本条件和高频技能等显性需求分析，逐步拓展至技能抽取、主题发现及关联关系挖掘，为岗位能力需求分析提供了基础。

1.2　大数据及相关专业人才培养研究

在大数据及相关专业人才培养方面，已有研究主要聚焦于数据管理类人才以及信息管理类人才等研究对象，通过挖掘招聘文本，从岗位职责、知识结构、专业技能及综合素质等方面分析岗位需求。朝乐门等[19]基于多个国家的招聘公告归纳了数据科学家的岗位职责与能力要求；周晓燕等[20]分析了国内数据管理人才的知识结构与岗位职责；薛洁等[21]对数据类岗位的人才需求特征及校招、社招差异进行了比较分析；邹凯等[22]探讨信息管理类专业技能、综合素质需求与人才培养改革之间的关系。聚焦大数据管理与应用专业，学者们多从培养方案、课程体系和培养模式等角度进行

研究。例如，赵星等[23]通过分析多所高校的培养方案，发现部分高校人才培养存在“重技术、轻管理”倾向；王晰巍等[24]提出以跨学科联合培养和产学协同育人为导向的培养模式，强调技术型、应用型与业务型人才的多层次培养；李兰兰等[25]搜集、整理国内84所高校大数据管理与应用专业的培养方案和课程体系，分析综合、理工、财经和师范4类院校在大数据管理与应用专业课程设置上的共性、特性及现存问题，并提出优化建议。

综上，现有研究已形成以内容分析和文本挖掘为主要路径的人才需求分析方法，并逐步由岗位基本条件和高频技能等显性需求分析拓展至技能抽取、主题发现及关联关系挖掘，同时围绕大数据及相关专业人才培养开展了较为丰富的研究。但现有研究对大数据管理与应用专业的关注仍显不足，尤其在岗位能力要素及其组合关系的系统分析、岗位能力需求向课程体系与实践教学映射等方面仍有待深化。基于此，本研究聚焦大数据管理与应用专业，运用主题分析、技能抽取与能力共现分析方法，探索从岗位能力需求解析到人才培养体系构建的实现路径，以完善该专业需求侧的实证研究，为专业人才培养优化提供实践依据。

2　研究设计

2.1　研究思路

本研究旨在回答三个核心问题：一是大数据管理与应用专业相关岗位需求呈现何种基本特征，以及从岗位需求中可以识别出怎样的能力主题；二是不同岗位类型呈现出怎样的能力组合以及存在何种差异；三是如何将岗位能力需求转化为可执行的人才培养优化策略。本研究以在线招聘文本为数据基础，遵循“需求侧—供给侧”的总体思路构建研究框架（见图1）。在需求侧阶段，首先开展数据收集与处理，从前程无忧网站采集大数据管理与应用专业相关岗位的招聘信息，对岗位职责、任职要求文本进行清洗等预处理，形成待处理数据集，然后结合招聘信息与专业文献构建招聘信息词典，用于辅助文本分词与专业词汇识别，再根据一致性与困惑度指标来确定最优主题数量。其次通过对学历需求、经验需求、地区分布、岗位类型等维度的统计分析，分析岗位的基本需求特征，并采用LDA模型识别岗位能力主题，通过技能抽取与同义归并形成岗位能力清单，结合能力共现关系识别岗位能力组合。在供给侧阶段，本研究从职业能力匹配、培养过程实施、培养成效反馈三个层面构建人才培养优化策略。

2.2　数据获取与词表构建

前程无忧网站的招聘信息覆盖行业广泛、岗位类型丰富，能反映当前就业市场对大数据相关人才的需求。本研究聚焦大数据管理与应用专业，通过文献调研并结合2位该领域专家的访谈意见，以“大数据工程师”“数据分析师”“数据运营”“数据产品经理”“数据管理”“数据仓库”“大数据开发”为关键词在前程无忧网站进行职位检索。本研究利用Python

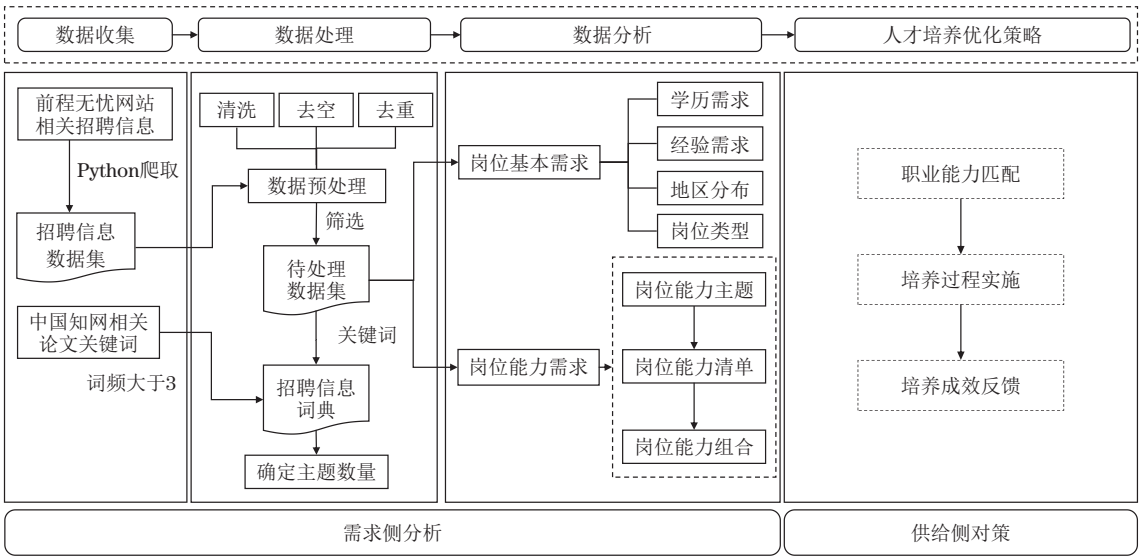

图1　研究框架

爬取招聘信息，提取职位名称、薪资情况、工作地点、工作经验、学历要求、福利待遇、岗位职责及任职要求等字段。随后，对获取的数据进行清洗等预处理，筛选标准为：①岗位职责或任职要求中涉及数据采集、存储、分析、治理、可视化或数据驱动决策等内容；②剔除岗位职责为销售、行政等专业关联度低的职位。最终获得3 390条有效招聘记录，数据采集时间为2025年2月12—17日。为提高主题分析的准确性，本研究结合招聘信息与专业文献构建领域词典。首先，招聘信息中的关键词能较好地反映岗位对专业和能力需求，经去重清洗，共得到999个关键词。其次，为提高技能词汇标准化程度，本研究引入专业知识关键词，考虑近几年的论文关键词可以降低早期研究专业术语体系尚未稳定带来的干扰，因此在中国知网以“大数据管理与应用”为关键词进行主题检索，对发表时间在2023年1月—2025年2月的论文进行关键词词频统计，从564篇论文中筛选出词频大于3的关键词131个。再次，加入前程无忧网站招聘信息中工作经验、能力、学历、经验等方面的关键词。最后，将以上关键词进行合并去重，得到1 160个关键词，构成本研究的招聘信息词典。

2.3 研究方法

2.3.1 LDA模型

LDA模型[26]作为一种无监督学习方法，能从文本中提取隐含主题，捕捉语义关系，适合处理大规模、多样化和复杂性的能力需求数据。本研究将招聘信息中的岗位职责和任职要求合并作为待分析文本，并利用前述构建的招聘信息词典对文本进行分词与专业词汇识别，在此基础上采用LDA模型进行主题分析。为提高主题识别结果的稳定性与可解释性，在建模前结合主题一致性与困惑度确定主题数量。

一致性分析用于衡量主题内词语的相关性。主要有点互信息（Pointwise Mutual Information，PMI）、归一化点互信息（Normalized Pointwise Mutual Information，NPMI）等基于词语的共现统计计算方法。目前广泛使用的是NPMI，其取值范围为[–1，1]，若NPMI值为1，则表示两个词语完全共现；若NPMI值为0，则表示两个词语独立；若NPMI值为–1，表示两个词语完全不共现。NPMI的计算方法如式1所示。

$$\mathrm{NPMI}\left(w_i,w_j\right)=\frac{\ln\dfrac{\mathrm{P}\left(w_i,w_j\right)}{\mathrm{P}\left(w_i\right)\cdot\mathrm{P}\left(w_j\right)}}{-\ln\mathrm{P}\left(w_i,w_j\right)} \quad (1)$$

其中，$\mathrm{P}\left(w_i\right)$是词语$w_i$在语料库中出现的概率，$\mathrm{P}\left(w_j\right)$是词语$w_j$在语料库中出现的概率，$\mathrm{P}\left(w_i,w_j\right)$是词语$w_i$和$w_j$共现的概率。

对于一个主题T，其一致性得分的计算方法如式（2）所示。

$$\mathrm{Coherence}\left(T\right)=\frac{2}{N\left(N-1\right)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{NPMI}\left(w_i,w_j\right) \quad (2)$$

其中，N是主题中词语的数量，w_i和w_j是主题中的词语对，主题一致性得分越高，说明主题内词语的语义相关性越强[27]，主题更易于理解和解释。

较低的困惑度表示模型对测试数据具有较好的预测性能，困惑度计算方法如式（3）所示。

$$\mathrm{Perplexity}\left(D_{\mathrm{test}}\right)=\exp\left(-\frac{\sum_{d=1}^{M}\sum_{w=1}^{N_d}\ln\mathrm{P}(w\mid d)}{\sum_{d=1}^{M}N_d}\right) \quad (3)$$

其中，D_{test}是测试集文档集合，M是测试文档的数量，N_d是文档d中的词语数量，$\mathrm{P}(w\mid d)$是模型预测的词语w在文档d中出现的概率，困惑度越低，表示模型对新数据的预测能力越强。结合招聘文本的语料规模、研究目的及主题可解释性，本研究将候选主题数量设置为1～15，并分别计算不同主题数量下的一致性与困惑度指标（见图2）。

依据不同主题数量对应的一致性得分（图2a）进行降序排序，选取前40%对应的主题数构成候选集合，分别是4、5、6、7、8、9。结合不同主题数量对应的困惑度（图2b），主题数为4、6、8时出现较明显的拐点。综合一致性与困惑度指标，两者交集为主题数4、6、8。考虑主题可解释性与相邻节点变化幅度，本研究选取主题数4作为最优主题数量。为保障结果稳定可靠，本研究多次重复运行模型进行验证，各主题的关键词与权重未发生明显漂移，一致性与困惑度指标平稳，表明模型结果具有可重复性与稳健性。

2.3.2 技能抽取与同义归并

为识别岗位能力的具体构成，本研究将招聘文本中的技能要求作为岗位能力的可观测要素，通过技能抽取与同义归并完成标准化处理。具体而言，在主题识

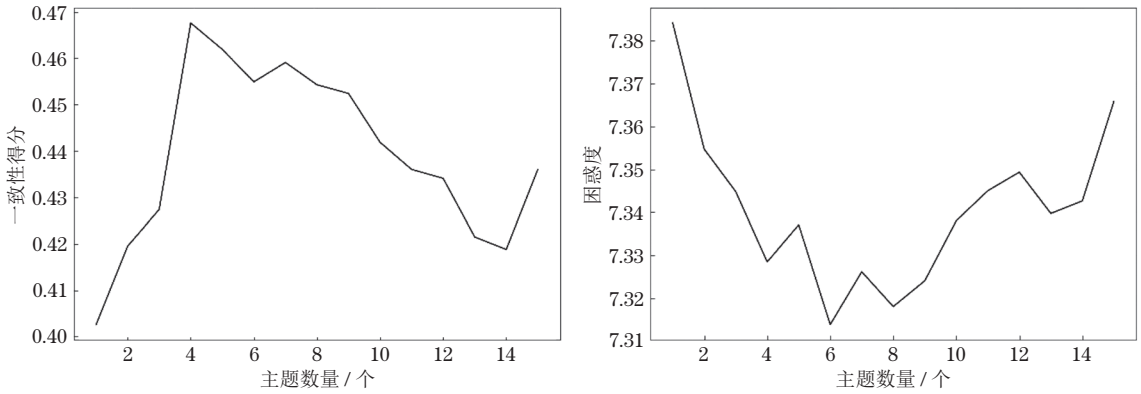

图2　不同主题数量的一致性得分和困惑度

别基础上，本研究借助大模型辅助开展细粒度技能抽取，将运用LDA模型提取出的4个主题簇内的岗位职责和任职要求的合并文本作为输入，设定统一提示指令为：“请对以下大数据岗位招聘的岗位职责与任职要求文本中，精准抽取专业技能、工具软件、平台框架、业务方法、职业素养五类实体；只输出标准化技能术语，剔除形容词、程度副词、无意义修饰词；不输出解释、不重复、不拆分语义，按条目罗列。”在抽取过程中，本研究对中英文混写、缩写及同义表达进行规范化处理，同时区分技能实体与任务描述；结合人工复核与一致性检验，对误抽取和低置信度技能项进行修正或剔除。在此基础上，本研究进一步对技能进行层级化整理，将高频稳定的技能作为一级节点，将具体工具、方法或应用场景作为二级节点，形成各主题簇的岗位能力清单。

2.3.3　能力组合识别

为识别岗位能力之间的共现关系及其形成的能力组合，本研究开展能力共现分析。将每条招聘记录对应的能力集合视为该岗位能力需求的表现，利用Python提取同一招聘记录内技能的共现信息，并运用Gephi构建能力共现网络；网络中节点表示具体技能，边表示技能之间的共现关系。为降低低频节点与弱关联边对网络结构解读造成的干扰，本研究对节点和边进行人工检验和过滤处理。在此基础上，本研究进一步凝练具有明确业务指向的稳定能力组合。将出现频次高、共现关系稳定，构成岗位通用能力底座的组合划分为基础能力组合；将对业务理解、工程实践和跨场景应用要求更高，更多集中于中高级岗位的组合划分为高级能力组合。

3　岗位基本需求

为把握大数据管理与应用相关岗位的基本需求特征，本研究对招聘信息中的学历需求、经验需求、地区分布及岗位类型进行统计分析。

从学历需求看，要求本科学历的岗位占比最高，达到68.75%；其次为大专（19.58%）和硕士（9.73%）。不同学历层次对应的岗位需求存在一定差异：硕士岗位更侧重算法研发与数据建模等高阶能力；本科岗位对数据分析等应用型能力的需求较为突出；大专岗位更多从事数据分析、数据采集等流程化基础工作。

从经验需求看，近半数岗位面向有工作经验者。要求3年以上工作经验的岗位占49.20%，“无需经验”的岗位占20.12%，明确面向应届生的岗位仅占1.86%。这表明当前就业市场既存在一定数量的基础性岗位，又更加重视具有项目经历和独立工作能力的中高级人才，岗位经验要求呈现较为明显的层次性。

从地区分布看，岗位需求主要集中于北京、上海、武汉、成都等核心城市和区域中心城市，并逐步向部分中西部城市、制造业城市扩散，整体呈现核心城市集聚与区域扩散的空间特征。

根据岗位关键词词云图（见图3）可将招聘岗位归纳为三类：一是以数据分析师、数据治理工程师、数据库管理员等为代表的数据类岗位，侧重数据采集、清洗、分析、可视化及数据治理；二是以数据开发工程师、算法工程师、软件工程师等为代表的工程研发类岗位，侧重数据平台开发、分布式计算、模型训练与部署；三是以产品经理、项目主管、数据运营等为代表的业务与管理类岗位，更强调业务理解、需求沟通、项目协作以及依托数支撑决策的能力。

图3　岗位关键词词云图

4　岗位能力需求分析

4.1　岗位能力主题

本研究采用LDA模型提取各主题的高权重特征词。结合大数据管理与应用专业的知识结构和岗位业务语境，对4个主题进行归纳命名，分别为：智能化项目管理与协作、数据驱动的业务优化分析、客户关系与销售运营管理、数据系统开发与数据库技术运用。各岗位能力主题的Top30特征词如表1所示。

表1　岗位能力主题及Top30特征词

岗位能力主题	特征词
智能化项目管理与协作	技术，项目，团队，模型，应用，业务，开发，系统，优化，AI，领域，专业，管理，设计，大数据，产品，需求，客户，研究，研发，算法，平台，Python，数据治理，合作，学习，服务，维护，落地，沟通
数据驱动的业务优化分析	业务，数据分析，团队，产品，专业，开发，设计，需求，运营，工具，优化，项目，系统，熟练，平台，沟通，管理，合作，SQL，客户，Python，技术，沟通能力，信息，销售，应用，流程，计算机，整理，数据库
客户关系与销售运营管理	项目，技术，客户，团队，产品，管理，需求，销售，业务，开发，应用，系统，工具，合作，质量，沟通，设计，专业，维护，优化，数据分析，流程，AI，Python，市场，熟练，有效，培训，理解
数据系统开发与数据库技术运用	数据库，技术，开发，专业，系统，应用，团队，业务，优化，模型，产品，项目，设计，需求，数据分析，工具，维护，平台，解决，Python，管理，SQL，Java，沟通，客户，算法，评估，理解，AI，基础

通过对特征词的分析可知，智能化项目管理与协作能力强调项目规划、任务执行、团队协作及跨部门沟通，并借助人工智能技术提升项目管理效率；数据驱动的业务优化分析能力侧重运用Python、SQL、统计分析及机器学习等工具和方法开展数据分析、业务建模与决策优化；客户关系与销售运营管理能力主要涉及CRM、用户画像、市场分析和运营管理，强调通过数据洞察客户需求并支撑销售与运营决策；数据系统开发与数据库技术运用能力以SQL、Python、Java、数据仓库等为基础，强调数据库操作、数据系统开发及数据基础设施建设。

4.2　岗位能力清单

本研究通过技能抽取与同义归并进一步构建各岗位能力主题对应的岗位能力清单。岗位能力主题—岗位能力清单映射如表2所示。

表2　岗位能力主题—岗位能力清单映射

岗位能力主题	岗位能力清单	能力类型覆盖
智能化项目管理与协作	需求分析、项目计划/里程碑、敏捷/迭代、协同工具、文档与汇报、跨部门沟通、AI辅助办公/自动化	软技能、方法、工具
数据驱动的业务优化分析	SQL、Python、统计分析、数据可视化、A/B测试、回归/分类/聚类、特征工程、指标体系与分析报告	工具、方法、业务
客户关系与销售运营管理	CRM、用户画像、市场/渠道分析、漏斗分析、指标看板、客户分群、运营策略评估、沟通协作与方案呈现	业务、工具、软技能
数据系统开发与数据库技术运用	数据库设计、SQL优化、抽取—转换—加载（Extract-Transform-Load，ETL）、数据仓库/数据湖、分布式计算、数据管道与调度、接口开发、数据安全与权限控制	平台、工具、工程

经分析发现，大数据管理与应用专业岗位能力需求体现为数据工程能力、数据分析能力、业务运营能力和项目协同能力4类能力的协同融合特征。这表明大数据管理与应用专业人才培养应围绕各能力模块构建课程群，加强工程实践、业务应用与综合素养培育，实现技术能力与管理能力、分析能力与业务能力的有机融合。

4.3　岗位能力组合

为规避人工划分能力组合带来的较强主观性，本研究采用Louvain模块度优化算法对能力共现网络进行主题簇划分。本研究在算法聚类得到若干初始主题簇的基础上，结合簇内技能节点的共现特征、语义关联以及大数据管理与应用岗位业务语境，剔除网络中的噪声节点，最终得到结构清晰、关联紧密的能力共现网络（见图4）。通过逐一解读各社区包含的核心技能要

素，发现网络聚类结果与前文LDA主题建模挖掘的能力模块语义高度契合，为实现两种研究结果的交叉印证、统一分析框架，本研究沿用一致的命名体系。

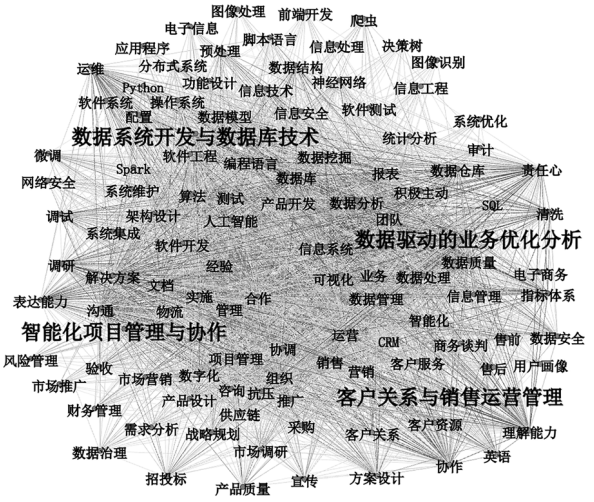

图4　能力共现网络

能力共现网络显示，各主题簇内技能之间具有较为稳定的共现关系，同时不同主题簇之间也存在一定的跨领域关联，表明岗位能力需求并非单一技能的独立要求，而是呈现多技能协同组合的特征。在此基础上，本研究结合技能共现关系、岗位职责及专业培养目标，对共现结果进行归纳，进一步提炼各主题簇内的岗位能力组合（见表3）。

分析结果表明，不同能力主题簇内部均存在若干高频能力组合。例如，在“数据驱动的业务优化分析”主题簇中，“SQL、Python、统计分析、可视化报表”构成基础分析能力组合，而“A/B测试、因果推断/回归、指标体系、业务解读”更常见于对业务理解要求较高的岗位。

5　人才培养优化策略

本研究基于前述岗位能力需求分析结果，以岗位能力需求与专业人才培养的有效衔接为出发点，从职业能力匹配、培养过程实施和培养成效反馈三个方面提出相应的优化策略。在职业能力匹配方面，高校可依据岗位能力主题及能力组合优化课程群；在培养过程方面，高校可深化项目制教学、推进真实任务实践，强化数据治理与合规教育，同时加强生成式人工智能应用能力培养；在培养成效方面，高校可通过学习成果评价，建立全程评价与动态调整机制。

5.1　优化岗位能力导向的课程群

大数据管理与应用专业课程体系应兼顾国家战略导向与市场岗位需求，立足现有专业课程体系，将岗位能力需求作为课程优化的重要参考，在服务数字经济发展和释放数据要素价值等国家战略方向的基础上，通过能力主题、课程群、训练任务、学习成果评价的对应关系，将岗位能力转化为可教学、可训练与可评价的培养目标。考虑到数字化转型和数据要素制度建设的发展趋势，尽管招聘信息中数据治理与合规相关岗位需求尚不突出，在专业建设层面仍应适当前置布局相关培养内容。因此，本研究在岗位能力分析基础上，构建岗位能力主题—课程群—项目任务—学习成果评价映射表（见表4）。

高校可在现有专业课程体系的基础上，参考表4所反映的岗位能力需求与课程教学之间的映射关系，对相关课程进行优化：一是结合专业培养目标和岗位能力需求，对照现有课程体系分析相关能力的课程支撑情况，识别需要补充或强化的内容；二是在保持专业知识体系完整性和课程设置基本稳定的基础上，加强相关课程之间的内容衔接与协同，并根据实际需要对课程内容更新或补充；三是结合相关课程设置项目制/实训，建立学习成果评价机制，通过分析报告、可运行脚本、可视化看板等成果形式考查学生对知识与技能的

表3　主题簇内的岗位能力组合

主题簇	基础能力组合	高级能力组合	对培养的直接指向
智能化项目管理与协作	需求分析、敏捷迭代、协同工具、文档交付	项目计划、风险管理、跨部门沟通、AI辅助办公	项目管理课程群；企业真实项目推进；形成“项目复盘与交付文档”作品集
数据驱动的业务优化分析	SQL、Python、统计分析、可视化报表	A/B测试、因果推断/回归、指标体系、业务解读	数据分析课程群；业务数据分析项目；形成“分析报告与看板”作品集
客户关系与销售运营管理	CRM、用户画像、客户分群、运营评估	渠道分析、漏斗分析、指标看板、策略建议	运营分析模块；营销/运营数据项目；形成“运营策略方案”作品集
数据系统开发与数据库技术运用	数据库设计、SQL优化、ETL、数仓/湖仓	Spark等分布式计算、调度编排、数据管道、权限控制	数据工程课程群；数据管道搭建项目；形成“管道代码与技术文档”作品集

表4　岗位能力主题—课程群—项目任务—学习成果评价映射表

能力主题（需求侧）	课程群（供给侧）	项目制/实训（训练侧）	学习成果评价（结果侧）
数据工程与系统开发	数据库原理、数据仓库/湖仓、分布式计算、数据工程化	数据管道搭建、湖仓一体项目、工程化代码规范	可运行SQL脚本、ER图、ETL流程图、部署文档、性能测试报告
业务分析与决策支持	统计分析、数据挖掘/机器学习、可视化与BI、指标体系与数据产品	真实业务数据分析报告、A/B实验或因果推断实践、仪表盘交付	分析报告、可视化看板、指标口径文档、数据集
智能化项目管理与协作	项目管理、数据治理导论、沟通与写作	跨角色协作项目、敏捷迭代、需求评审与汇报	画像标签体系、策略报告、效果预估表、业务落地建议
客户运营与场景落地	数据驱动运营、用户研究、行业案例分析	客户画像与运营策略设计、CRM/运营数据案例复盘	需求文档、项目里程碑、复盘报告、评审记录
数据治理与合规安全	数据治理、数据标准与质量、数据安全与隐私保护	数据资产盘点、质量规则配置、合规评估报告撰写	治理方案、质量规则清单、合规评估报告、数据字典

综合运用能力。

5.2　深化项目制实践与真实任务驱动

从招聘岗位经验需求看，3年及以上中高级岗位占比高，应届生专属岗位偏少，说明用人单位普遍重视候选人的项目经验与实际交付能力，这一结论与商务智能[9]、信息管理与信息系统[10]等领域岗位招聘需求研究的发现基本一致。结合本研究岗位能力组合分析结果，数据管道搭建、业务数据分析、客户画像与运营策略设计、需求分析与项目协作等任务，均体现出多项能力协同运用的特点。因此，实践教学应进一步突出真实岗位任务驱动模式，依据招聘文本中的岗位职责和能力组合提炼典型任务，并将其转化为课程或跨课程项目。教学过程可以引入企业真实或准真实数据、业务场景和工作要求，按照任务提出、实施分析、成果交付、评价反馈的过程组织教学，引导学生完成从需求理解、数据处理到方案形成和成果交付的完整任务，并以分析报告、可视化看板、技术文档、策略方案等作为评价成果，提升学生的综合实践与项目交付能力。

5.3　强化数据治理与合规教育

技能抽取结果显示，招聘文本中已零星出现数据质量、权限管理、隐私合规等方面的能力要求，结合数字化转型与数据政策发展背景，完善数据治理与合规能力培养具有一定必要性。随着数据要素价值释放和数据基础制度建设推进，数据治理、数据安全与隐私保护逐渐成为数据人才培养的重要内容[28]。因此，开设大数据管理与应用专业的高校可在数据工程相关课程中融入数据质量控制、元数据管理和权限配置等内容；同时依托数据治理、数据安全与个人信息保护相关课程，将法律法规、行业规范与制度要求融入专业教学实践。具体可以通过数据资产盘点、数据质量规则配置、权限方案设计和合规评估等实践任务，将相关知识转化为数据字典、质量规则清单、合规评估报告等可操作的实践成果，提升学生在数据治理与规范应用方面的能力。

5.4　加强生成式人工智能应用能力培养

随着生成式人工智能技术的快速发展，其在数据分析、知识检索和辅助决策等场景中的应用不断深化[29]；同时人工智能等关键词在岗位需求描述中频繁出现，因此高校有必要在课程体系中加强对学生生成式人工智能应用能力的培养。一方面，可在数据分析、业务决策和项目实践相关课程中引入人工智能辅助分析、自动化特征工程与智能报表生成等典型应用场景，重点培养提示设计、结果校验、偏差识别和可解释表达能力，使学生能够在具体业务任务中合理使用生成式人工智能工具；另一方面，可将人工智能伦理与治理相关内容纳入教学，引导学生认识模型幻觉、算法偏见与责任边界模糊等潜在风险，提升学生技术判断、风险识别与规范应用能力。

5.5　健立全程评价与动态调整机制

由于岗位能力需求会随技术发展和产业环境变化而调整，人才培养策略也需要根据需求变化持续优化。在全程评价方面，高校可围绕培养目标、课程实施、项目训练与能力达成情况建立全程评价机制，将课程考核、项目成果、作品集、企业评价和毕业去向反馈等作

为培养质量的重要证据，并引入企业导师参与课程共建、项目评审与作品集评价，形成从学习产出到岗位胜任的对照标准。在动态调整方面，高校可建立基于招聘数据的年度或半年度需求追踪制度，将能力组合作为课程内容调整的重要依据，根据岗位需求变化及时补充、调整相关课程内容与实践任务。

6　结语

本研究基于在线招聘文本，对大数据管理与应用专业的岗位能力需求进行系统识别，并据此提出人才培养的优化路径。研究表明：岗位能力并非单一技术的简单叠加，而是在数据基础与工程化、分析建模与业务洞察、数据治理与安全合规等方面形成协同结构；不同岗位类型与层级对应差异化的能力组合，高层级与高薪岗位更强调业务理解、数据治理与跨部门协作，基础岗位则更侧重工具栈与工程实施能力；大数据管理与应用专业人才培养应从“课程清单导向”转向“岗位能力需求导向”，依托项目制教学，形成任务、产出、学习成果评价的全链条评价体系。

相较于已有研究，本研究主要贡献如下：一是将岗位需求识别进一步延伸至能力组合构建，通过能力共现分析揭示能力的协同特征；二是聚焦大数据管理与应用专业，将各类岗位能力凝练为能力组合，并建立能力组合与课程群的映射关系；三是构建由岗位需求、能力组合到课程体系的培养优化路径，为高校专业人才培养策略的动态调整提供可量化、可验证的依据。

本研究仍存在一定局限性。一是样本主要来自单一招聘平台，后续可融合多平台及不同类型机构的招聘信息，并结合时间序列分析，刻画岗位能力需求的动态演化；二是本研究主要从企业需求侧探讨专业培养优化路径，未来可结合高校课程体系、毕业生就业反馈及企业访谈等提升研究结论的科学性与推广价值。

参考文献

[1] 刘业政，孙见山，姜元春，等. 大数据的价值发现：4C模型[J]. 管理世界，2020，36（2）：129–138，223.

[2] 杨丽，徐绪堪，李一铭. 面向大数据管理与应用专业的数据素养教育研究[J]. 情报理论与实践，2020，43（10）：45–49，18.

[3] 教育部关于公布2017年度普通高等学校本科专业备案和审批结果的通知[EB/OL]. [2026–05–02]. http://www.moe.gov.cn/srcsite/A08/moe_1034/s4930/201803/t20180321_330874.html.

[4] 邓胜利，汪璠，夏苏迪. 大数据管理与应用本科专业学位教育发展状况调查报告（2023）[J]. 图书情报知识，2024，41（1）：69–79，101.

[5] 邓胜利，袁梦，汪璠. 大数据管理与应用本科专业学位教育发展状况调查报告（2024）[J]. 图书情报知识，2026，43（2）：158–167.

[6] 国务院关于印发促进大数据发展行动纲要的通知[EB/OL]. [2026–05–02]. https://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.

[7] 工业和信息化部关于印发“十四五”大数据产业发展规划的通知[EB/OL]. [2026–05–10]. https://wap.miit.gov.cn/jgsj/xxjsfzs/wjfb/art/2021/art_a7a3ea67af254d13a6a07d5a098e727f.html.

[8] 国家职业分类大典修订工作委员会. 中华人民共和国职业分类大典：2022年版[M]. 北京：中国劳动社会保障出版社，2022.

[9] 胡忠义，李雅，吴江，等. 基于招聘信息的商务智能人才需求分析与启示[J]. 信息资源管理学报，2019，9（3）：111–118，129.

[10] 陈明红，张倩琳，韩静. 信息管理与信息系统专业人才招聘需求分析及培养启示[J]. 图书馆学研究，2021（20）：9–20.

[11] 王伟. 就业市场对金融科技人才的需求：基于2 448条招聘广告的文本分析[J]. 财会月刊，2021（18）：110–115.

[12] 贾东琴，刘阳禾，卢嘉怡. 我国公共图书馆员核心技能框架构建探究：基于岗位招聘需求的文本分析[J]. 图书馆，2026（4）：42–52.

[13] 李尚昊，郝琦. 内容分析与文本挖掘在信息分析应用中的比较研究[J]. 图书馆学研究，2015（23）：37–42，69.

[14] 袁毅，陶鑫琪，李瑾萱，等. 基于招聘文本实体挖掘的人才供需分析：以人工智能领域为例[J]. 图书情报工作，2022，66（14）：101–118.

[15] 王东波，胡昊天，周鑫，等. 基于深度学习的数据科学招聘实体自动抽取及分析研究[J]. 图书情报工作，2018，62（13）：64–73.

[16] 俞琰，陈磊，姜金德，等. 网络招聘文本技能信息自动抽取研究[J]. 图书情报工作，2019，63（13）：105–113.

[17] 崔文波，张涛，许鑫. 情报学领域典型文本主题聚类算法比较研究[J]. 情报理论与实践，2025，48（7）：199–207.

[18] 岳铁骐，傅友斐，徐健. 基于招聘广告的岗位人才需求分析框架构建与实证研究[J]. 数据分析与知识发现，2022，6（Z1）：151–166.

[19] 朝乐门，肖纪文，王解东. 数据科学家：岗位职责、能力要求与人才培养[J]. 中国图书馆学报，2021，47（3）：100–112.

[20] 周晓燕，尹亚丽. 基于国内市场需求的大数据管理人才知识结

构分析[J]. 情报科学，2017，35（1）：29–34.

[21] 薛洁，王娴，籍艳丽. 大数据时代数据类岗位人才需求特征分析[J]. 黑龙江高教研究，2022，40（10）：119–124.

[22] 邹凯，罗文鑫，蒋知义，等. 大数据时代信息管理类人才社会需求的调研与分析[J]. 情报理论与实践，2022，45（11）：54–60.

[23] 赵星，俞晓婷，万玲玉. 新文科背景下大数据管理与应用专业培养特征的内容分析[J]. 图书与情报，2020（6）：26–34，92.

[24] 王晰巍，李玥琪，贾若男，等. 新文科背景下大数据管理与应用专业人才培养模式[J]. 图书情报工作，2021，65（17）：45–56.

[25] 李兰兰，焦建玲，杨冉冉，等. 新文科背景下大数据管理与应用专业课程体系建设研究[J]. 图书馆学研究，2023（10）：2–12.

[26] BLEI D M，NG A Y，JORDAN M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research，2003，3：993–1022.

[27] ARUN R，SURESH V，VENI MADHAVAN C E，et al. On finding the natural number of topics with Latent Dirichlet Allocation：some observations[C]//Advances in Knowledge Discovery & Data Mining：14th Pacific-Asia Conference. Berlin：Springer，2010：391–402.

[28] 叶忠明，赵璐. “管工结合”模式下管理类人才数据治理技能培养的实践探索[J]. 档案管理，2023（1）：104–105，108.

[29] 张颖怡，王玉琢，张恒，等. 生成式AI背景下多方需求驱动的课程内容体系设计：以“信息分析”课程为例[J]. 图书馆论坛，2025，45（11）：83–94.

作者简介

孙倬，男，博士，副研究员，博士生导师，研究方向：信息行为、人本人工智能、数据安全及治理。
李嘉欣，女，硕士研究生，研究方向：数据要素、文本挖掘。
鲁晓明，女，硕士，副教授，硕士生导师，通信作者，研究方向：信息组织、知识图谱，E-mail：xmlu@zzu.edu.cn。
杨瑞仙，女，博士，教授，博士生导师，研究方向：信息计量、数据要素、社会网络。

Analysis of Job Competency Requirements and Talent Cultivation Optimization Strategies for Big Data Management and Application Major: Based on Text Mining of Recruitment Information

SUN Zhuo[1,2,3] LI JiaXin[1] LU XiaoMing[1,2] YANG RuiXian[1,2,3]
(1. School of Information Management, Zhengzhou University, Zhengzhou 450001, P. R. China; 2. Henan Data Governance Research Centre, Zhengzhou 450001, P. R. China; 3. Research Center for “Double World-Class Project” of Henan Province, Zhengzhou 450001, P. R. China)

Abstract: Against the backdrop of deepening digital transformation and the integration of industry and education, the Big Data Management and Application major faces a mismatch between the supply of talent and job market demand. This study selected online job postings from the 51job website to conduct a statistical analysis of basic job requirements, including educational background, experience, job type, and geographic distribution. It employed the Latent Dirichlet Allocation (LDA) model to identify competency themes and combined skill extraction with competency co-occurrence analysis to characterize the competency requirements for these positions. The study found that basic job requirements exhibit four key characteristics: the educational threshold is primarily a bachelor’s degree; nearly half of the positions require prior work experience; positions are mainly distributed in core cities and regional central cities; and job types can be categorized into data-related, engineering and R&D, and business management roles. Job competency requirements primarily revolve around four competency dimensions: intelligent project management and collaboration; data-driven business optimization and analysis; customer relationship and sales operations management; and data system development and database technology application. Furthermore, there are significant differences in competency requirements across different job types. Based on this demand-side analysis of job competency requirements, this study proposes strategies for optimizing talent development in the field of big data management and application, providing a practical foundation for professional talent cultivation and contributing to the development of the national big data industry.

Keywords: Big Data Management and Application; Recruitment Information; Text Mining; Competency Combination; Demand Analysis; Talent Development