doi: 10.20239/j.issn.1671-9492.2026.09.014

基于赋存状态多维语义特征向量化的锂矿石相似性评价方法研究

杨佳伟[1,2],张康辉[1,2],高希宇[1,3],张晓亮[1,3],余瑾彦[1,2],朱阳戈[1,3]

(1. 矿冶科技集团有限公司,北京 100160;
2. 矿冶过程智能优化制造全国重点实验室,北京 102628;
3. 矿物加工科学与技术全国重点实验室,北京 102628)

摘 要: 随着矿产资源品位下降与矿石类型日益复杂,矿石赋存状态差异显著、矿物学特征维度多样,传统经验驱动的选矿工艺设计方法已难以满足资源高效利用的需求。针对复杂矿石赋存特征难以统一表达和定量比较的问题,本文提出基于赋存状态多维语义特征向量化的矿石相似性评价方法。首先,构建包含矿石类型、化学成分、矿物组成、嵌布特征及解离度等多维属性的结构化特征体系,并引入语义特征向量化方法,实现复杂矿物学信息的数字化与语义化表达。其次,基于词频-逆文本频率指数(Term frequency-inverse document frequency,TF-IDF)特征向量化与余弦相似度算法,建立矿石赋存特征的相似性评价模型,并通过人工设定的特征重要性权重调节机制提升关键赋存属性的区分能力。通过对待分析矿石样本与典型矿石样本库的语义相似度对比,实现不同赋存状态矿石在特征层面的相似性度量与筛选。研究结果表明,该方法可实现复杂矿石赋存特征的统一表征,能够从矿物学特征维度量化矿石间的相似程度;需注意矿石赋存特征相似并不直接等价于选矿工艺及分选指标一致,本研究可为后续开展选矿工艺知识迁移、工艺候选方案匹配与智能决策研究奠定数据与方法基础。

关键词: 矿石赋存状态;多维语义特征;特征向量化;相似度分析

中图分类号: TD913;TD954 **文献标志码:** A **文章编号:** 1671-9492(2026)09-0122-10

A Method for Ore Similarity Evaluation Based on Multidimensional Semantic Feature Vectorization of Occurrence States

Yang Jiawei[1,2], Zhang Kanghui[1,2], Gao Xiyu[1,3], Zhang Xiaoliang[1,3], Yu Jinyan[1,2], Zhu Yangge[1,3]

(1. BGRIMM Technology Group, Beijing 100160, China;
2. State Key Laboratory of Mining and Metallurgy Process Manufacturing Optimization, Beijing 102628, China;
3. State Key Laboratory of Mineral Processing, Beijing 102628, China)

Abstract: With the continuous decline in mineral resource grades and the increasing complexity of ore types, significant differences in ore occurrence states and the high dimensionality of mineralogical characteristics have made traditional experience-driven beneficiation process design methods insufficient to meet the requirements of efficient resource utilization. To address the difficulty of unified representation and quantitative comparison of complex ore occurrence features, this study proposes an ore similarity evaluation method based on multidimensional semantic feature vectorization of occurrence states. Firstly, a structured feature system is constructed by integrating multiple attributes, including ore type, chemical composition, mineral assemblage, textural characteristics, and degree of liberation. Semantic feature vectorization techniques are then introduced to achieve digitalized and semantic representations of complex mineralogical information. Subsequently, an ore similarity evaluation model is established using term frequency-inverse document frequency (TF-IDF) feature vectorization and cosine similarity, and an empirically-defined feature importance weighting adjustment mechanism is incorporated to enhance the discriminative

收稿日期: 2026-01-04
基金项目: 地球深部探测与矿产资源勘查国家科技重大专项(2024ZD1002000)
作者简介: 杨佳伟(1982—),男,内蒙乌兰察布人,博士,高级工程师,主要从事矿物加工智能化与过程优化控制研究。
通信作者: 张康辉(1995—),男,河南周口人,博士,工程师,主要从事矿物加工过程优化控制研究。

capability of key occurrence attributes. By comparing the semantic similarity between target ore samples and a database of representative ore samples, similarity measurement and screening of ores with different occurrence states at the feature level are realized. The results demonstrate that the proposed method can realize unified representation of complex ore occurrence features and quantitatively evaluate ore similarity from mineralogical perspectives. It should be noted that similar mineralogical occurrence features do not guarantee consistent beneficiation performance. This work provides a data and methodological foundation for subsequent research on beneficiation process knowledge transfer, candidate-scheme matching, and intelligent decision-making.

Key words: ore occurrence state; multidimensional semantic feature; feature vectorization; similarity analysis

随着新能源、新材料及高端制造业的快速发展，铁、铜、锂、钨等关键金属矿产在能源储备、新能源电池、高性能材料及战略性产业中的需求持续增长[1-2]，对战略性矿产资源的综合评价、精细化开发以及智能化利用提出了更高要求。矿石赋存状态作为决定资源可选性和开发潜力的重要基础，其精细化认知与定量表征已成为选矿智能化研究的重要前提。然而，我国矿产资源赋存条件复杂多样，不同矿床在矿石类型、矿物组合、颗粒大小及矿物嵌布特征等方面存在显著差异，导致矿石加工难度和选矿过程复杂度随之增加[3-4]。矿石赋存特征呈现出多源、多尺度和强非线性的特点，传统矿物学分析结果往往以定性描述或分散指标形式存在，难以实现多维信息的统一表达与综合比较。目前，矿物学分析与选矿工艺设计较大程度上依赖人工经验，缺乏系统的结构化数据和可计算的特征表达方式[5]。这种以经验为主导的方法难以对矿石赋存状态的多维特征进行准确量化和相似性评价，限制了不同矿石样本之间的横向对比与知识复用能力，难以为后续选矿工艺研究提供可迁移、可推理的特征基础。在此背景下，亟需建立一种能够对矿石赋存状态多维特征进行结构化建模、语义化表达与定量相似性评价的方法体系，为复杂矿石特征的客观比较与分类提供技术支撑，并为后续开展选矿工艺知识迁移与智能决策研究奠定基础[6-7]。

在选矿工艺设计实践中，矿石赋存状态、矿物颗粒尺寸、矿物间嵌布关系及化学成分等多维因素对工艺流程适应性和冶金效率具有重要影响[8-10]。传统经验驱动的方法在处理复杂矿石时存在明显局限：一方面，矿石特征信息表达零散，关键赋存特征难以实现统一量化；另一方面，不同矿石样本之间缺乏有效的度量机制。为改善矿石特征信息的组织与表达方式，研究者尝试构建矿物学结构化数据库，对矿石类型、粒度、嵌布特征等多维属性进行系统建模，并结合统计分析与数据挖掘方法开展矿石特性研究。然而，不同特征维度之间的权重关系缺乏统一建模方法，现有研究多采用经验赋值方式确定维度权重，易造成评价结果主观性，影响评价结果的稳定性与可靠性。在矿物特征量化与分析方面，国内外研究利用矿物颗粒图像识别、扫描电镜(SEM)、X射线衍射(XRD)等技术提取矿物组成、粒度及嵌布特征，并通过数字图像处理与统计建模实现矿石微观特征的定量表达，为矿石性质分析提供了重要数据基础[11-13]。同时，一些研究建立了矿石解离度、嵌布部分等角度出发，建立矿石特征与选矿行为之间的关联关系。随着机器学习与数据驱动方法的发展，部分研究开始尝试将矿物学特征与数据分析模型相结合，实现矿石性质的分类、识别与预测[14-17]。此外，也有研究引入相似性分析思想，对不同矿石样本在特征层面进行对比，为工艺知识迁移和案例复用提供参考[18-19]。尽管上述研究在矿石特征提取与分析方面取得了一定进展，但总体来看，仍存在以下不足：其一，矿石赋存状态多维特征缺乏统一、结构化且具备语义信息的表达方式，难以实现跨样本的有效比较；其二，不同特征维度在矿石相似性评价中的贡献关系尚未得到系统建模，现有研究多采用经验赋值方式确定维度权重，易造成评价结果主观性，影响评价结果的稳定性与可靠性。

针对上述问题，本文提出了一种基于赋存状态多维语义特征向量化的矿石相似性评价方法。该方法通过构建涵盖矿石类型、化学成分、矿物组成、嵌布特征及解离度等多维属性的结构化特征模型，并引入语义特征向量化技术，实现矿物学信息的数字化与语义化表达。在此基础上，结合词频-逆文本频率指数(Term frequency-inverse document frequency，TF-IDF)[20]特征向量化与余弦相似度算法，建立矿石赋存特征的相似性评价模型，并通过特征权重调节机制增强关键赋存特征的区分能力，从而实现不同赋存状态矿石在特征层面的相似性度量与筛选，可为选矿候选工艺方案推荐、工艺知识迁移与智能决策研究

提供数据与方法支撑，矿石特征相似仅作为工艺参考依据，不能完全替代选矿试验验证。

1 数据与方法

1.1 数据结构化处理

本研究基于19份完整的锂矿选矿试验报告，系统提取了矿石特征的10个维度，包括：矿石类型及矿床信息、主要化学成分分析、锂物相分析、矿物组成及相对含量、矿石结构构造、主要目的矿物的嵌布特征、主要锂矿物粒度组成、不同磨矿细度下锂矿物解离度测定、赋存状态及矿物学因素分析。为确保数据的完整性与可计算性，所有原始数据均以Excel表格形式汇总，并按照各维度建立统一的数据字段规范，部分数据示例见表1～4。

在数据处理过程中，首先对Excel表格中的文本信息、数值数据及试验结果进行清洗与标准化，包括缺失值填充、异常值剔除以及单位统一处理。随后，通过设计自动化程序对Excel数据进行解析，将各维度信息转换为结构化数据格式，并生成对应的JSON文件(图1)。每个JSON文件包含矿石样本编号及10个维度的字段信息，实现数据的标准化、可访问和可计算化，为后续多维语义特征向量化和智能工艺匹配提供基础数据支撑。

在数据清洗过程中，部分选矿报告存在个别矿物学维度字段缺失的情况。对于缺失的特征维度，在后续相似度计算中将该维度权重置为0，不参与加权相似度计算，避免无效数据干扰整体评价结果。

1.2 多维语义特征向量化

在完成矿石特征数据结构化处理后，为实现复杂矿石赋存状态的数字化表达与相似度计算，本研究对10个维度的矿石特征进行了语义层面的向量化建模。通过引入自然语言处理(Natural language processing，NLP)与特征编码技术，将文本型与数值型矿物学信息统一映射至高维语义向量空间，实现了矿石赋存状态的多维语义化表示。首先，对JSON格式数据中的文本信息(如矿石类型、结构构造及矿物学因素分析等)进行中文分词处理，以提取关键语义特征。为提高分词的准确性与专业性，结合领域知识建立专用矿物学词典，并采用基于前缀词典的最大正向匹配算法与动态规划算法相结合的分词策略。

1)基于前缀词典的最大正向匹配算法

依据原始语料构建前缀词典，存储所有可能词语及其对应的词频信息。对于输入句子，从左至右逐字扫描，每次尽可能匹配词典中最长的词语，实现最大正向匹配。通过建立词语频率、词性及前缀树结构，可高效生成所有可能成词组合的有向无环图(Directed acyclic graph，DAG)，其中节点表示词语起始位置，边表示可能的词语结束位置。随后，在DAG结构上应用动态规划算法计算各路径概率，并选择概率乘积最大的路径作为最优切分方案，其概率定义如下：

$$P(S)=\prod_i p(w_i)=\prod_i \frac{\mathrm{FREQ}(w_i)}{\sum \mathrm{FREQ}(w_i)} \tag{1}$$

其中，$p(w_i)$表示词语w_i的出现概率，$\mathrm{FREQ}(w_i)$为词频，该方法能够有效平衡词语长度与出现的概率，使分词结果更加符合矿物学领域语义特征。

2)基于隐马尔科夫模型的动态规划算法

针对前缀词典未覆盖的领域新词或专有名

表1 矿石主要化学成分分析

Table 1 Main chemical composition analysis of the ore /%

成分	Li_2O	Rb_2O	SiO_2	K_2O	Na_2O	Al_2O_3	CaO	MgO	Fe_2O_3	CaF_2
含量	0.226	0.17	74.56	4.03	2.82	12.54	0.36	0.15	1.24	0.71

注：含量<0.1的未列出。

表2 矿石锂物相分析

Table 2 Lithium phase analysis of the ore /%

锂相别	云母中锂	离子吸附锂	其他锂	合计
含量	0.221	0.001	0.002	0.224
分布率	98.66	0.45	0.89	100.0

表3 矿石矿物组成及相对含量

Table 3 Mineral composition and relative content of the ore /%

矿物名称	铁锂云母	白云母	铌钽铁矿	石英	钠长石	钾长石	黄玉	高岭石	萤石	磁铁矿
含量	6.14	1.82	0.02	42.95	23.61	16.77	4.82	0.92	0.66	0.11

注：含量<0.1的未列出。

表 4　矿石主要锂矿物的粒度组成

Table 4　Particle size composition of main lithium minerals of the ore　/%

粒级/mm	云母矿物集合体	
	含量	累计
+3.5	3.51	3.51
−3.500+3.000	4.46	7.97
−3.000+2.500	6.30	14.27
−2.500+2.000	7.03	21.3
−2.000+1.651	8.07	29.37
−1.651+1.168	12.36	41.73
−1 168+0.833	11.76	53.49
−0.833+0.589	8.52	62.01
−0.589+0.417	11.87	73.88
−0.417+0.295	8.12	82.00
−0.295+0.208	6.41	88.41
−0.208+0.147	4.35	92.76
−0.147+0.104	3.37	96.13
−0.104+0.074	2.03	98.16
−0.074+0.043	1.38	99.54
−0.043+0.020	0.31	99.85
−0.020+0.015	0.080	99.93
−0.015+0.010	0.050	99.98
−0.010	0.020	100.0

词，引入隐马尔科夫模型(Hidden Markov model，HMM)[21]进行补充识别。首先计算HMM的参数，包括每个字符作为词语开头、中间或结尾的概率；然后构建状态转移网络，将输入句子映射为由字符状态组成的转移图；最后采用动态规划算法(如Viterbi算法)搜索概率最大的状态路径。该路径对应的字符组合即为最优分词结果，从而实现对新词及专业术语的识别与补全。

通过上述两种算法的结合，实现了基于统计与概率推理的语义化分词结果，为后续TF-IDF向量化及相似度计算提供了高精度的词语粒度输入。此方法在保证矿物学专业术语完整性的同时，显著提升了文本信息在高维语义空间中的可区分性与表达一致性，为矿石赋存状态特征的量化与语义建模奠定了技术基础。

3)TF-IDF特征向量化与特征融合

在完成文本语义切分与关键特征提取后，为实现矿石语义信息的量化表示，本研究采用词频-逆文档频率(TF-IDF)算法对分词结果进行特征向量化处理。TF-IDF是一种基于统计的文本权重计算方法，用于衡量词语在语料中的重要性，其核心思想是：某一词语在特定文档中出现频率越高，同时在所有文档中出现频率越低，则该词语对该文档的区分能力越强。TF-IDF的计算公式如下：

$$\mathrm{TF\text{-}IDF}(t,d)=\mathrm{TF}(t,d)\times\mathrm{IDF}(t) \tag{2}$$

$$\mathrm{IDF}(t)=\log\frac{N}{n_t+1} \tag{3}$$

其中，$\mathrm{TF}(t,d)$表示词语t在文档d中的出现频率，$\mathrm{IDF}(t)$为逆文档频率，N为文档总数，n_t为包含词语t的文档数。通过该方法，可计算出每个词语在样本文档中的加权权重，从而生成具有统计意义的高维语义向量。

在本研究中，每份选矿报告被视为一个文档，报告中提取的矿石特征信息作为语料输入。经TF-IDF处理后，得到的向量可以反映矿石赋存状态文本特征的词频权重分布，进而体现矿石类型间的部分语义差异；受TF-IDF词袋模型本身限制，该方法难以识别矿物学领域的同义词、近义表述，对文本深层语义的表征能力存在固有局限。为充分整合矿物学特征中


```
{
  "矿石基本信息": {
    "矿石类型": "锂多金属矿中含有价元素锂、钽、铌、铍等，矿石中锂绝大部分以独立矿物铁锂云母的形式存在",
    "矿床详细": "属于碱性长石型花岗岩型锂云母矿床"
  },
  "主要化学成分": {
  "锂物相分析": {
    "锂物相分析结果": {
    "钽物相分析结果": {
    "铌物相分析结果": {
  },
  "矿物组成及相对含量": {
  "矿石的结构构造": "矿石为二长伟晶岩，主要呈由结晶作用所形成的粗粒结构。 粗粒结构：矿石中的铁锂云母、长石和石英等矿物一般呈粗粒结构分布",
  "主要目的矿物的嵌布特征": {
    "铁锂云母电子探针成分分析结果_电子探针": [
    "白云母电子探针成分分析结果_电子探针": [
  },
  "主要锂矿物的粒度组成": {
    "矿石中云母矿物的粒度组成": {
  },
  "不同磨矿细度下锂矿物的解离度测定": {
    "不同磨矿细度下锂矿物的解离度测定": {
    "各磨矿细度产品中云母矿物集合体的嵌布粒度": {
  },
  "赋存状态": {
    "Li2O在不同矿物中的分布": {
  },
  "矿物学因素分析":
  "矿石中锂主要以类质同象的形式存在于铁锂云母和白云母中，两种矿物嵌布粒度较粗，嵌布关系紧密，且浮游性能接近，通过浮选可以做到较好的回收。少量铁锂云母以细粒形式嵌布于脉石矿物中，较难与脉石矿物单体解离，将以连生体或脉石矿物包体的形式损失。提升锂回收率的途径主要是强化连生体及微细粒云母的回收。影响最终锂精矿 Li2O 含量提高的因素有以下几点：
  (1)矿石中铁锂云母含量为6.14%，其中 Li2O
  的平均含量为3.58%。但同时，矿石中有1.82%的白云母，其平均Li2O含量为0.23%，并且白云母与铁锂云母嵌布关系紧、可浮性相近，在选矿流程中走向一致。因此，若将矿石中的铁锂云母和白云母全部浮选进入锂精矿，可获得的锂精矿最高理论品位为2.81%；(2)矿石中的云母矿物集合体粒度较粗，+0.3mm粒级部分占82.00%，适当粗磨即可以达到较好的解离效果。但存在少量铁锂云母呈细片状嵌布于脉石矿物中，在磨矿过程中较难与脉石矿物单体解离，将以裸露贫连生体或脉石矿物包体的形式损失于尾矿中；(3)矿石中存在较多的云母和少量高岭石，在磨矿过程中易泥化，从而恶化选矿环境，影响锂的回收。矿石中铌、钽绝大部分以铌钽铁矿形式存在，微量以细晶石形式存在，另有微量的铌、钽以类质同象的形式存在。铌钽铁矿的嵌布粒度较细，在磨矿过程中较难与脉石矿物单体解离，将以贫连生体或脉石矿物包体的形式损失。提升铌、钽回收率的途径主要是强化钽铌矿物的解离及微细粒钽铌矿物的选矿回收。"
}
```


图 1　标准化 JSON 文件示例

Fig. 1　Example of a standardized JSON file

的文本信息与数值信息，本研究进一步进行了特征融合处理。首先，将数值型特征(包括主要化学成分、矿物组成、粒度组成及解离度等)进行归一化与标准化处理，消除不同量纲间的影响；随后，采用向量直接拼接的方式将数值型特征向量与文本语义特征向量进行融合，形成统一的多维特征表示向量，该拼接融合方式存在模态不均衡风险，高维稀疏文本向量有可能压制部分数值特征的贡献。此外，为体现各特征维度对选矿工艺适配的重要性差异，引入特征权重调整机制。本研究基于矿物学机理与工程经验人为赋予各维度差异化权重，尚未采用数据驱动方法实现权重自动寻优；其中化学成分、矿物组成与嵌布特征被设定为高权重特征，而结构构造与矿物学因素分析则作为辅助因子参与模型计算。最终，经过加权融合后的特征向量能够更全面地表征矿石赋存状态的语义信息与理化特征，为后续的相似度分析与工艺匹配提供高质量的输入基础。

1.3 相似度计算与工艺匹配机制

在完成矿石赋存状态多维语义特征的向量化表达后，为实现新矿石与典型样本间的智能匹配与工艺推荐，本文构建了基于余弦相似度(Cosine similarity)的赋存状态语义相似度计算模型，并结合特征权重调整机制建立了选矿工艺智能匹配体系。

1)余弦相似度计算

余弦相似度是一种常用于高维向量空间中度量样本间相似度的指标，通过计算两个特征向量夹角的余弦值来反映其语义或属性上的相似程度。其计算公式如下：

$$\mathrm{Sim}(A,B)=\frac{A\times B}{\|A\|\times\|B\|}=\frac{\sum_{i=1}^{n}a_ib_i}{\sqrt{\sum_{i=1}^{n}a_i^2}\times\sqrt{\sum_{i=1}^{n}b_i^2}} \quad (4)$$

其中，A与B分别为待匹配矿石样本与典型矿石样本的多维特征向量，a_i与b_i为第i个特征维度的取值。相似度值$\mathrm{Sim}(A,B)$介于[0,1]之间，值越大表示两个样本在语义特征空间中的相似程度越高。

为提高模型在复杂矿石样本下的适应性，本研究引入特征权重系数向量$W=\{w_1,w_2,\cdots,w_n\}$，对不同维度特征进行差异化加权处理。综合考虑矿石赋存机理及工艺敏感性，权重设计遵循以下原则。

(a)主控特征高权重化：如矿物组成、嵌布特征、粒度组成等直接影响选矿工艺选择的因素赋予较高权重；

(b)辅助特征低权重化：如矿物学因素、结构构造等描述性特征赋予较低权重；

(c)可调节性：在实际应用中，可通过历史工艺匹配准确率对权重进行迭代优化。

加权余弦相似度计算公式为：

$$\mathrm{Sim}_w(A,B)=\frac{A\times B}{\|A\|\times\|B\|}=\frac{\sum_{i=1}^{n}w_i\times a_ib_i}{\sqrt{\sum_{i=1}^{n}w_i\times a_i^2}\times\sqrt{\sum_{i=1}^{n}w_i\times b_i^2}} \quad (5)$$

2)工艺匹配与推荐机制

在相似度计算完成后，本文构建了工艺样本知识库，该库包含19个典型矿石的工艺流程及对应特征向量信息，涵盖磨矿、分级、浮选、磁选、重选等典型选矿单元操作。针对输入的新矿石样本，系统首先计算其与知识库中各典型样本的加权相似度，并根据相似度得分进行排序。本研究人为设定相似度划分阈值：若最高相似度样本$\mathrm{Sim}_w\geq0.85$，则输出该典型样本对应的工艺流程作为候选参考方案；若相似度介于0.6～0.85，则系统依据相似度权重对多个样本的工艺特征进行加权融合，输出综合候选参考方案。当相似度低于0.60时，系统将提示需补充试验数据或人工验证，以防误匹配。上述阈值基于本研究19份锂矿样本的经验划定，普适性有待更多样本进一步校验，不同矿种、不同矿石类型下合理阈值可能存在差异。

此外，本文引入知识驱动反馈机制作为框架构想：理论上可在每次工艺匹配后，将选矿试验验证结果回传至数据库，用于动态更新特征权重与样本库，从而实现“数据-模型-决策”的闭环优化；该闭环迭代机制尚未开展实际工程运行验证。

2 结果与讨论分析

2.1 锂矿样本语义相似性分析结果

为验证所提出方法在锂矿样品间语义差异识别及工艺适配中的有效性，本文选取湖南省某锂多金属矿与四川某锂多金属矿两份完整选矿试验报告作为对比样本，基于10个维度的结构化数据开展语义相似度分析，各维度语义相似度结果见图2。

从总体结果来看，两组锂矿样品的加权相似度为0.338 6，属于低相似度水平，表明其矿物学特征及赋存条件存在显著差异(表5)，具体分析如下。

1)高相似度维度(相似度>0.6)主要集中于“矿物组成及相对含量”(0.614)和“主要目的矿物的嵌布特征”(0.657)。两地锂矿在主要含锂矿物类型和嵌布粒度分布方面表现出一定一致性，均以锂辉石与锂云母为主要赋存相，反映出其在矿物组合和载体结构上的共性。

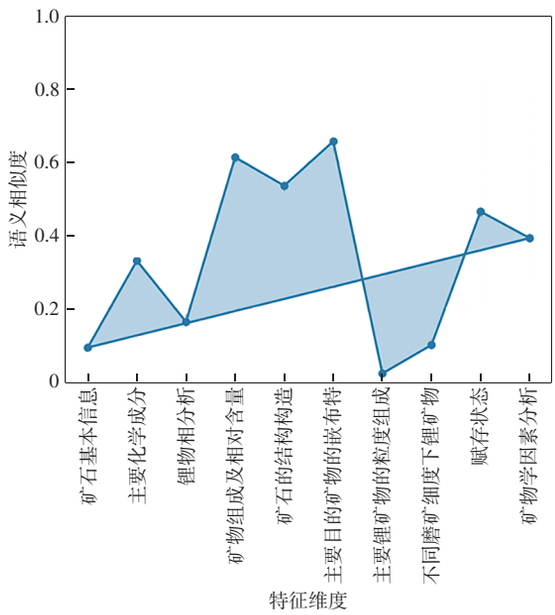


图 2　各维度语义相似度结果

Fig. 2　Results of semantic similarity in each dimension

2) 中等相似度维度(0.4～0.6)包括“矿石结构构造”(0.536)与“赋存状态”(0.465)。该结果说明两矿在成岩构造和矿化环境上具有一定相似性，可能受相近成矿作用或岩浆演化阶段控制，但区域地质背景与矿化演化路径仍存在差异。

3) 低相似度维度(相似度＜0.4)占比较高，尤其是“主要锂矿物的粒度组成”(0.022)与“不同磨矿细度下锂矿物的解离度”(0.101)差异显著，表明两矿在磨矿细度响应规律及解离行为方面存在明显差异，反映出其选矿过程中对粒度控制的敏感性不一致。此外，“锂物相分析”(0.164)和“主要化学成分”(0.331)相似度较低，说明两矿在含锂相类型及伴生元素组合上存在本质差别。

综合来看，湖南与四川锂矿虽在矿物组成及赋存状态方面具有一定相似性，但在粒度分布、物相组成及选矿行为等关键维度上差异显著，导致总体语义相似度偏低(0.339)。本多维语义特征向量化模型可识别矿石矿物学层面的多维差异特征，可为锂矿选矿工艺差异化设计提供参考依据；矿石特征相似度仅反映矿物学赋存层面差异，不能直接等价于分选行为、选矿指标的相似。

2.2　多维语义特征下的锂矿样本相似性分析

为进一步揭示语义相似性模型在不同矿物学特征维度上的区分能力，本文基于10个结构化维度对湖南某锂多金属矿(样品1)与四川某锂多金属矿(样品2)进行了多维语义特征比对分析，见图3。模型同时考虑了文本语义相似度、数值特征差异度及专业语义调整权重，最终得到加权总体相似度为0.338 6，表明两地矿石在语义层面存在显著差异。

1) 矿石基本信息

该维度的最终相似度为0.093 4，属于低相似等级。主要差异体现在矿石类型和成矿背景上：样品1为碱性长石型花岗岩型锂云母矿，锂主要以铁锂云母形式存在；样品2为花岗伟晶岩型锂辉石矿，成矿环境与锂赋存相显著不同。两矿类型匹配度低(类型相似度0.024)，说明两地矿石的成因模式差异较大。

2) 主要化学成分

两锂矿中均包含主要氧化物元素(如SiO_2、Al_2O_3、K_2O、Na_2O等)，共同成分覆盖率达到64%。然而关键含锂组分差异显著：样品1的Li_2O含量为0.226%，而样品2达1.28%，相对差异0.823。尽管主要脉石组分(SiO_2、Al_2O_3)相似度较高(＞0.8)，但含铌、钽、铍等伴生元素含量差异极大，特别是Ta_2O_5与Nb_2O_5相似度低于0.001，导致整体化学相似度较低。该维度的含量相关系数为0.594 5，表明主要成分间具有

表 5　锂矿样本语义相似度结果汇总

Table 5　Summary of semantic similarity results of lithium samples

序号	特征维度	相似度值	语义相似性等级	备注
1	矿石基本信息	0.093 4	低	区域成矿背景差异显著
2	主要化学成分	0.330 8	低	锂、铍、铌含量差异大
3	锂物相分析	0.163 6	低	含锂相矿物类型不同
4	矿物组成及相对含量	0.613 6	中	主矿物类型相似，含量差异
5	矿石的结构构造	0.536 1	中	构造特征部分一致
6	主要目的矿物的嵌布特征	0.657 4	中	嵌布粒度分布相近
7	主要锂矿物的粒度组成	0.022 5	低	粒度差异显著
8	不同磨矿细度下锂矿物的解离度	0.100 5	低	解离规律不一致
9	赋存状态	0.465 3	中	赋存形态差异较小
10	矿物学因素分析	0.392 7	低	矿物共生关系不同
总体加权相似度	—	0.338 6	低	—

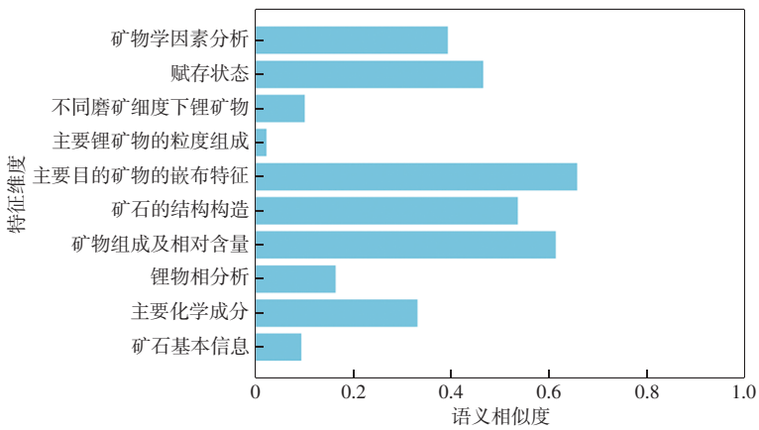


图3 各维度相似性对比

Fig. 3 Comparison of similarity in each dimension

一定一致性,但关键矿物元素显著不同。

3)锂物相分析

样品1中锂主要分布于云母矿物(98.66%)中,而样品2报告未提供可比的锂物相数据。两者在“锂分布率”和“锂含量”指标上缺乏有效对应项,因此该维度相似度计算结果无效,在整体加权相似度计算中该维度权重置0,不参与总得分运算。

4)矿物组成及相对含量

该维度调整后相似度为0.613 6,属于中等相似水平。两矿均以石英、钾长石、钠长石及铁锂云母为主要矿物,共同矿物覆盖率为0.67。其中铁锂云母含量一致(6.14%),石英含量差异12.6%,Na长石相似度最高(0.866),白云母含量差异显著(相似度0.178)。总体高含量矿物相似度为0.796,表明两矿在主要脉石结构上具有较强一致性。

5)矿石结构构造

样品1报告描述为“二长伟晶岩,呈粗粒结构”;样品2未提供结构性文字信息。尽管文本语义匹配度低,但模型基于结构关键词(“粗粒”、“结构”)识别出局部相似语义,因此判定为中等相似水平,反映出两矿在岩体成因上存在一定共性。

6)主要目的矿物的嵌布特征

该维度相似度最高,为0.657 4。两矿报告均包含电子探针与能谱数据,样品1测点68个、样品2测点62个。共同分析元素多达15种(Al_2O_3、FeO、K_2O、SiO_2等),成分相似度0.552。尽管样品1以铁锂云母为主要载体,而样品2以锂辉石为主,但在嵌布粒度及元素组合上表现出较高的相似性,说明两者在选矿可磨性与矿物颗粒分布规律上具有一定一致特征。

7)主要锂矿物的粒度组成

相似度仅0.022 5,为所有维度中最低。样品1主要涉及云母类矿物粒度分布,而样品2描述锂辉石与铌钽铁矿粒度组成。两者矿物类型完全不重叠,矿物覆盖率为0,粒度特征不可比,体现出两地矿石在嵌布解离特征上的本质差异。

8)不同磨矿细度下锂矿物的解离度

该维度的相似度为0.100 5。尽管两报告均提供了磨矿试验数据,但样品1和样品2在锂矿物单体解离率、连生体含量和解离度模型上差异显著,该维度反映出两矿在选矿工艺响应曲线上的差异性较大。

9)赋存状态

该维度相似度为0.465 3。样品1中Li_2O主要赋存于白云母与铁锂云母中,而样品2主要赋存于锂辉石、锂绿泥石及磷锰锂矿。共同赋存形式仅“白云母”一项,Li_2O含量相似度为0.391。说明两地锂的赋存载体差异显著,但存在部分相似的层状硅酸盐赋存特征。

10)矿物学因素分析

该维度相似度为0.392 7。样品1中锂主要以类质同象形式赋存于铁锂云母和白云母中;样品2则以独立矿物锂辉石为主。样品1提出“需强化微细粒云母的回收”,样品2则强调“应加强对白云母的抑制”。两者在选矿优化目标上相反,语义相似度仅0.388,说明其工艺优化思路差异明显。

2.3 跨维度语义相关性分析

为进一步揭示不同矿物学特征维度之间的语义耦合规律,对10个语义维度进行了跨维度相关性计算,结果见图4。由图4可知,各维度之间相关系数普遍较低,表明不同矿物学属性在语义空间中具有较强的独立性与异质性特征。这一结果从侧面说明,不同维度主要刻画矿石赋存状态的不同方面,信息重叠程度较低,具备进行多维特征融合的合理性与必要性。

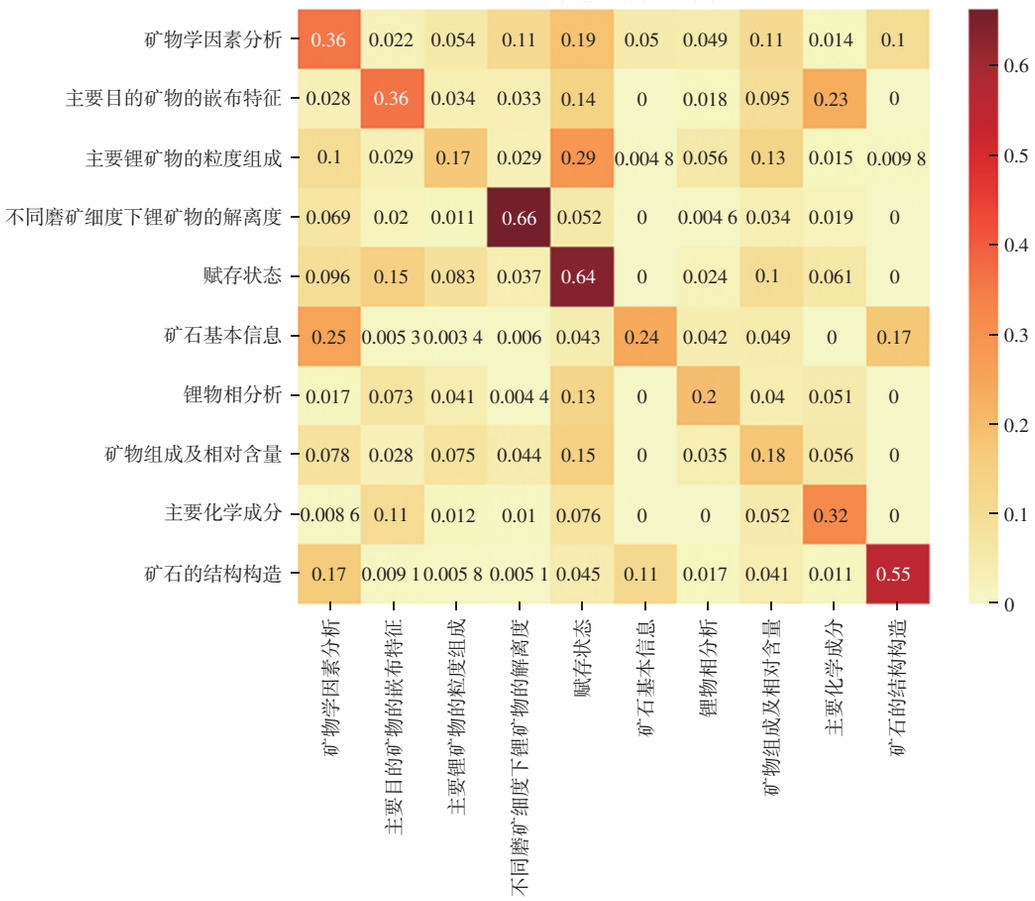


图 4　跨维度相似度热力图

Fig. 4　Cross-dimensional similarity heatmap

与此同时，在部分关键维度之间仍可观察到一定程度的相关性，反映了矿石结构特征、化学组成与选矿行为之间存在潜在的内在联系。

1）“赋存状态”与“主要目的矿物的嵌布特征”呈现中等相关（r=0.150），说明矿物的空间赋存方式在一定程度上受控于目的矿物的嵌布粒度与组合形式；同时，“矿物组成及相对含量”与“赋存状态”的相关系数为0.147，进一步印证了矿物组成对赋存形态的结构约束作用。

2）“矿石的结构构造”与“矿物学因素分析”的相关系数达到0.166，表明宏观结构特征在一定程度上反映了矿物学微观因素的综合效应；“矿石基本信息”与“矿石结构构造”之间的相关度（r=0.174）亦较高，暗示地质背景与构造特征在矿石物质组织中具有一致性控制作用。

3）在选矿行为相关维度中，“主要锂矿物的粒度组成”与“赋存状态”的相关系数为0.288，是全矩阵中最高的跨维度相关值。这表明矿物粒度特征对锂矿物的赋存状态具有显著影响，粒度分布差异可能导致不同矿石在磨矿解离与分选过程中的可选性差异。此外，“矿物学因素分析”与“不同磨矿细度下锂矿物的解离度测定”呈低至中等相关（r=0.106），反映出矿物组合复杂性与解离行为之间的非线性耦合关系。

计算得到的相关系数整体偏低，本部分未开展显著性检验，多数仅体现微弱相关趋势，仅能反映维度间潜在关联；这表明矿石的多维语义特征在语义空间中呈现出多层级、弱耦合的复杂结构，可为后续开展矿石类型智能聚类、选矿工艺参数预测提供初步参考。需要说明，本研究样本库仅包含19份锂矿选矿报告，样本数量有限，且全部为锂矿资源，模型泛化能力受到数据集约束；同时本研究未划分独立测试集，缺少交叉验证与消融试验定量对比不同模型方案的效果，后续可扩充样本、补充消融试验进一步验证模型性能。

3　结论

1）多维语义特征向量化可实现矿石赋存状态的数字化表征：通过构建矿石类型、化学成分、矿物组成、

嵌布特征及解离度等多维属性的结构化特征模型，并引入语义特征向量化方法，复杂矿物学信息得以数字化和语义化表达，为不同矿石样本在矿物学特征层面的量化比较提供了统一且可计算的表征方式。

2)不同矿区矿石在多维特征上差异明显且具有部分共性：相似性评价显示，不同矿区锂矿在矿物组成、赋存状态、粒度特征及磨矿解离度等关键维度存在显著差异，整体加权相似度较低(<0.4)，反映了伴生元素类型、锂相分布及潜在选矿行为的差异。与此同时，“主要目的矿物的嵌布特征”和“矿物组成及相对含量”等维度表现出一定共性(相似度>0.6)，表明主要锂矿物类型及其粒度分布在不同矿区存在一定相似特征。跨维度相关性分析得到部分维度间较低的相关系数，提示矿物粒度与赋存状态之间存在潜在关联趋势，可为多矿区矿石的分类与比较提供可量化依据。

3)本研究提出的多维语义相似性分析方法能够从矿物学赋存层面揭示不同矿石的共性与差异，可为多矿区矿石比较、差异化研究及大规模矿石样品的相似性度量提供分析工具，为后续知识驱动的矿产资源评价与智能化决策研究奠定数据与方法基础；应当注意，矿石赋存特征相似并不等同于选矿分选效果一致，工艺推荐结果仅作为候选参考，仍需选矿试验进一步验证。

4)受试验条件与数据获取成本限制，矿石样本数量有限，未设置独立外部测试集，对方法的统计稳健性与泛化能力存在一定影响；TF-IDF词袋模型对矿物学文本深层语义、专业同义词识别能力有限；多维特征权重主要基于矿物学机理认知与工程经验设定，尚缺乏数据驱动的动态自适应优化机制；文本-数值特征采用简单向量拼接融合，存在模态不均衡风险；工艺匹配阈值与闭环反馈机制也有待更多工程样本校验。后续将扩充样本规模，开展消融试验与阈值敏感性分析，引入数据驱动的权重学习方法，并推广至铁、铜等矿种，以提升方法通用性与工程适用性。

创新点：

1)采用语义向量化方法实现锂矿复杂矿物学特征的定量表达，开展矿石数字化对比分析。

2)通过跨维度相似性分析，揭示矿石各特征之间潜在的相关变化趋势。

3)基于语义的矿石相似性评价可为选矿方案筛选提供候选参考依据。

参考文献

[1] 俞柏. 战略性矿产资源供需形势分析与判断[J]. 中国金属通报，2025(8)：11-13.
Yu Bai. Analysis and judgment of supply and demand situation of strategic mineral resources[J]. China Metal Bulletin，2025(8)：11-13.

[2] 陈从喜. 战略性矿产资源开发利用研究进展与展望[J]. 中国矿业，2025，34(2)：4-19.
Chen Congxi. Progress and prospects in the development and utilization of strategic mineral resources[J]. China Mining Magazine，2025，34(2)：4-19.

[3] 李鹏，刘艳，唐晨宇，等. 中国高科技矿产资源供应安全现状与保障路径[J]. 工程科学学报，2025，47(11)：1985-1998.
Li Peng，Liu Yan，Tang Chenyu，et al. Current status and safeguarding pathways of supply security for high-tech mineral resources in China[J]. Chinese Journal of Engineering，2025，47(11)：1985-1998.

[4] 王登红. 关键矿产的研究意义、矿种厘定、资源属性、找矿进展、存在问题及主攻方向[J]. 地质学报，2019，93(6)：1189-1209.
Wang Denghong. Research significance，mineral classification，resource attributes，exploration progress，existing problems，and principal directions of critical minerals[J]. Acta Geologica Sinica，2019，93(6)：1189-1209.

[5] Zhironkina O，Zhironkin S. Technological and intellectual transition to mining 4.0：a review[J]. Energies，2023，16(5)：1427.

[6] 苏妤芸，苏秀珠，刘全军，等. 新疆某难处理金矿的金赋存状态研究[J]. 有色金属(选矿部分)，2025(10)：20-27.
Su Yuyun，Su Xiuzhu，Liu Quanjun，et al. Study on the occurrence state of gold in a refractory gold ore from Xinjiang[J]. Nonferrous Metals (Mineral Processing Section)，2025(10)：20-27.

[7] 孙传尧，周俊武. 流程工业选矿过程智能优化制造发展战略[J]. 有色金属(选矿部分)，2019(5)：1-5.
Sun Chuanyao，Zhou Junwu. Development strategy of intelligent optimization manufacturing for process-industry mineral processing[J]. Nonferrous Metals (Mineral Processing Section)，2019(5)：1-5.

[8] 吕燕，周家云，惠博，等. 吉林某铌钽多金属矿石中元素的赋存状态及对选矿工艺的影响[J]. 矿产综合利用，2019(2)：83-88.
Lv Yan，Zhou Jiayun，Hui Bo，et al. Occurrence state of elements in a Nb-Ta polymetallic ore from Jilin and its

influence on beneficiation technology[J]. Multipurpose Utilization of Mineral Resources, 2019(2): 83-88.

[9] 叶超，罗政. 某低品位铌矿石工艺矿物学特征及可选性研究[J]. 中国矿山工程，2021，50(6)：87-91，96.
Ye Chao, Luo Zheng. Study on technological mineralogical characteristics and processability of a low-grade niobium ore[J]. China Mine Engineering, 2021, 50(6): 87-91, 96.

[10] 张愿宁，郭园园，逯文辉，等. 甘肃敦煌五一山钒矿矿石特征及钒赋存状态[J]. 甘肃科技，2025，41(7)：37-43.
Zhang Yuanning, Guo Yuanyuan, Lu Wenhui, et al. Ore characteristics and vanadium occurrence state of the Wuyishan vanadium deposit in Dunhuang, Gansu[J]. Gansu Science and Technology, 2025, 41(7): 37-43.

[11] 朱汇派，费光春，谭华，等. 扫描电镜-能谱分析在四川西部某锂矿石铌钽元素赋存状态研究上的应用[J]. 有色金属(选矿部分)，2023(5)：16-21，41.
Zhu Huipai, Fei Guangchun, Tan Hua, et al. Application of SEM-EDS to the study of Nb-Ta element occurrence in a lithium ore from western Sichuan[J]. Nonferrous Metals (Mineral Processing Section), 2023(5): 16-21, 41.

[12] 汤化明，郎增瑞，王宏铭，等. 基于深度学习的金属矿物镜下嵌布粒度检测方法[J]. 地质科技通报，2024，43(3)：351-358.
Tang Huaming, Lang Zengrui, Wang Hongming, et al. Deep-learning-based method for measuring dissemination grain size of metallic minerals under microscope[J]. Bulletin of Geological Science and Technology, 2024, 43(3): 351-358.

[13] 蔡明明，李光胜，朱幸福，等. 基于金矿岩芯的自动工艺矿物学[J]. 矿产综合利用，2024，45(3)：174-178.
Cai Mingming, Li Guangsheng, Zhu Xingfu, et al. Automated process mineralogy based on gold-ore drill core[J]. Multipurpose Utilization of Mineral Resources, 2024, 45(3): 174-178.

[14] Yang Jun, Sun Chuanyao, Zhou Junwu, et al. Knowledge-inference-based intelligent decision making for nonferrous metal mineral-processing flowsheet design[J]. Minerals, 2025, 15(4): 374. DOI: 10.3390/min15040374.

[15] 王智宇，王达，邱昆峰，等. 机器学习在矿物岩石地球化学大数据挖掘中的应用与展望[J]. 成都理工大学学报(自然科学版)，2025，52(5)：844-858.
Wang Zhiyu, Wang Da, Qiu Kunfeng, et al. Application and prospect of machine learning in big-data mining of mineralogy and geochemistry[J]. Journal of Chengdu University of Technology (Natural Science Edition), 2025, 52(5): 844-858.

[16] 田妞，黄宋魏，和丽芳，等. 深度学习在选矿领域的应用进展[J]. 化工矿物与加工，2023，52(8)：69-74，82.
Tian Niu, Huang Songwei, He Lifang, et al. Application progress of deep learning in the field of mineral processing[J]. Industrial Minerals & Processing, 2023, 52(8): 69-74, 82.

[17] Wang Qiang, Zhang Kai, Zou Guobing, et al. Advances in deep learning-based ore particle size detection: a review of methods, challenges, and trends[J]. MetaResource, 2025, 2(2): 83-104.

[18] 于潇然，徐泉，初延刚，等. 基于相似性检测的选矿指标趋势监控系统[J]. 控制工程，2020，27(增刊1)：8.
Yu Xiaoran, Xu Quan, Chu Yangang, et al. A trend-monitoring system for mineral-processing indices based on similarity detection[J]. Control Engineering, 2020, 27(S1): 8.

[19] 于潇然. 选矿生产指标趋势监控与分析系统的设计与开发[D]. 沈阳：东北大学，2020.
Yu Xiaoran. Design and development of a trend-monitoring and analysis system for mineral-processing production indices[D]. Shenyang: Northeastern University, 2020.

[20] Al-Obaydy W N I, Hashim H A, Najm Y A, et al. Document classification using term frequency-inverse document frequency and K-means clustering[J]. Indonesian Journal of Electrical Engineering and Computer Science, 2022, 27(3): 1517-1524.

[21] Tang Xiao. A latent hidden Markov model for process data[J]. Psychometrika, 2024, 89(1): 205-240.

（本文编辑　刘水红）