

基于 LightGBM 的盐湖锂浓度遥感反演研究

——以西藏扎布耶盐湖北湖为例

刘婷玥¹⁾, 代晶晶²⁾, 赵元艺²⁾, 田淑芳¹⁾, 叶传永²⁾

1) 中国地质大学(北京)地球科学与资源学院, 北京, 100083;

2) 中国地质科学院矿产资源研究所, 北京, 100037

内容提要:青藏高原盐湖中的锂、硼等矿产资源具有极高的经济价值。锂、硼等矿产资源的含量及其空间分布是盐湖观测的重点,对于盐湖资源的开发利用具有重要的指导意义。利用遥感技术开展盐湖观测可以克服传统观测站观测空间上数据不连续、费时费力等缺点,而机器学习等人工智能算法可以快速高效地挖掘遥感数据信息,因此本文基于 Landsat-8 遥感影像数据,利用 LightGBM 算法开展西藏扎布耶盐湖北湖锂浓度遥感反演。经过采样点波谱数据的获取、锂浓度 LightGBM 回归模型的构建、盐湖锂浓度反演一系列的实验步骤,最终的模型评价结果显示平均相对误差为 0.053530925,均方根误差为 10.2869,卡方为 0.867,模型与实测数据拟合程度较高。反演结果表明:整个北湖中,锂浓度最高的是东南部的水域,最低的是中西部的水域;河流和秋里南木泉水群的汇入使得附近的水域锂浓度降低;对于本身位于中部锂浓度低值区的钙华岛来说,岛内泉水的汇入使得附近的水域锂浓度有所升高,这一结果与实际情况较为吻合。通过本文的反演研究,证明了 LightGBM 机器学习算法用于快速反演盐湖锂浓度的可行性和精确性,同时为其它盐湖矿产资源遥感反演提供了技术启示,也为后续的盐湖资源量评估奠定了基础。

关键词:青藏高原;扎布耶盐湖;锂浓度;遥感反演;LightGBM

青藏高原分布着地球上海拔最高,数量最多和面积最大的内陆湖群,是中国湖泊分布最密集的地区之一,其中很多湖泊富集多种矿产资源,具有极高的经济价值(Lu Shanlong et al., 2016; Yang Kehan et al., 2017; Lei Li, 2019)。锂、硼等矿产资源的含量及其空间分布是青藏高原盐湖观测的重点内容,对于盐湖资源的提取具有重要的指导意义。从 20 世纪末开始,中国地质科学院在青藏高原盐湖周围陆续设立观测站,连续观测了近 30 年的盐湖气象、水文、水化学等方面变化,为盐湖的研究提供了宝贵的数据资料(Qi Wen et al., 2006a, 2006b; Kong Weigang et al., 2017)。但是观测站观测存在着一定的局限性,与其相比,遥感观测则从更宏观的角度观测盐湖,不仅能够提供研究对象不同时间

的空间分布状况,弥补观测站数据空间上不连续的缺陷,还能节省大量的人力、物力和财力(Zhang Chungui et al., 2013; Pan Xuepeng et al., 2015; Liu Fan et al., 2018; Fan Jiazhi et al., 2021),因此遥感技术为盐湖观测提供了一种良好的技术手段。

盐湖锂等矿产资源的遥感反演主要包括两类方法,一类是基于主成分分析、相关分析、盐指数和线性回归的反演方法,另一类是基于 BP (Back Propagation)神经网络、随机森林等机器学习算法的反演方法。Tian Shufang et al. (2005)和 Xu Wenning et al. (2017)利用主成分分析、相关分析等方法,对浓度高、成分复杂的扎布耶盐湖盐度、锂含量进行定量研究,建立了主成分与锂含量数据之间

注:本文为中国地质调查项目(编号 DD20190167)、中央级公益性科研院所基本科研业务费专项资金(编号 KK2012)及中央级公益性科研院所基本科研业务费专项资金(编号 KK2102)联合资助的成果。

收稿日期:2021-04-17;改回日期:2021-04-26;网络发表日期:2021-06-16;责任编辑:李曼。

作者简介:刘婷玥,女,1997 年生。硕士。地质工程专业。E-mail: 3394998821@qq.com。通讯作者:代晶晶,女,1982 年生。研究员,博导。遥感地质专业。E-mail: daijingjing863@sina.com。

引用本文:刘婷玥,代晶晶,赵元艺,田淑芳,叶传永. 2021. 基于 LightGBM 的盐湖锂浓度遥感反演研究——以西藏扎布耶盐湖北湖为例. 地质学报, 95(7): 2249~2256. doi:10.19762/j.cnki.dizhixuebao.2021222.
Liu Tingyue, Dai Jingjing, Zhao Yuanyi, Tian Shufang, Ye Chuanyong. 2021. Remote sensing inversion of lithium concentration in salt lake using LightGBM: a case study of northern Zabuye salt lake in Tibet. Acta Geologica Sinica, 95(7): 2249~2256.

的回归方程;Wang Junhu et al. (2015)构建了多维矿化度估测指数,并运用比值和主成分分析法估算了湖水的矿化信息;Wang Dejun et al. (2018)利用波段组合方法建立盐指数,再结合主成分分析对盐湖中的锂离子含量进行分离,完成了盐湖中锂离子含量遥感预测;Zhou Yamin et al. (2016)利用 BP 神经网络确定了最优波段组合并构建了盐湖矿物离子含量的反演模型;Wang Dejun(2019)利用随机森林、BP 神经网络、PCA 变换三种反演模型进行阿里萨罗盐湖的锂浓度反演,通过比较反演精度,得出了随机森林效果最好的结论。

LightGBM 作为一个 GBDT(梯度提升决策树, Gradient Boosting Decision Tree)算法框架,被广泛应用于各种回归预测中,大量实验证明,LightGBM 与传统的梯度提升决策树、随机森林、支持向量等算法相比,具有更高的精确度和更快的预测速度(Fan Dexiang, 2018; Sun Yue, 2019; Xie Yong et al., 2019; Pan Zhenbao et al., 2021)。西藏扎布耶盐湖是中国最大的锂资源矿,也是世界上三个超百万吨的大型盐湖之一,因此本文尝试以具有代表性的西藏扎布耶盐湖北湖为例,采用 LightGBM 机器学习方法,开展锂浓度遥感反演研究。本文旨在提供一种更精确快速的盐湖矿物遥感反演方法,为后续的盐湖资源量评估及开采利用提供技术支撑。

1 研究区概况

扎布耶盐湖是世界上唯一一个天然结晶碳酸锂的盐湖,地处青藏高原腹地,冈底斯山脉西段北麓,地理坐标东经 $83^{\circ}57'10''\sim 84^{\circ}15'08''$,北纬 $31^{\circ}27'10''\sim 31^{\circ}34'30''$ 。湖区隶属高原亚寒带气候区,具有气温低、温差大、辐射强、降水少、蒸发多的特点(Qi Wen et al., 2006b)。盐湖分为南湖和北湖。南湖固液并存,水深 2 m,北湖为卤水湖,水深数米。湖盆为封闭内流盆地,受断裂控制呈南北方向延伸,接受河水、大气降水和地下水补给(Tian Shufang et al., 2005)。湖区内河流、泉水众多,河流主要有浪门嘎曲、罗具藏布、桑目旧曲、甲布曲和泉水河,泉水主要分布在湖中部的钙华岛和湖北部的秋里南木,这些河流和泉水中含有大量的锂、硼等资源。区内出露的地层主要有石炭系、二叠系、白垩系、古近系、新近系和第四系(图 1)。河水中的锂、硼主要来源于第四纪沉积物、地表风化物及地下岩石,泉水中的锂主要来自于与深循环地下水接触的围岩(Liu Xifang et al., 2007)。与经过日晒蒸发的湖水相

比,河水、泉水的锂、硼浓度仍然较低。扎布耶湖水中的锂浓度在 1 g/L 以上,能形成天然碳酸锂沉积,非常有利于锂的提取。1982 年和 1984 年的盐湖综合调查显示,固液相锂资源量为 747 万 t,硼资源为 2458 万 t(Liu Xifang et al., 1999)。

2 数据与方法

常规的插值分析按照距离远近进行值的推算,而本文的锂浓度反演与插值分析不同的是,本文方法基于盐湖卤水中锂的波谱特性开展。地物波谱中往往包含着丰富的地物信息,不同锂浓度的卤水会在波谱上表现出不同的辐射值,因此能够更真实地反映盐湖中锂资源的分布状况。通过学习不同锂浓度下的波谱特征就能够得到辐射值与锂浓度之间的映射关系,从而进行锂浓度的预测和反演。基于 LightGBM 的盐湖锂浓度反演技术路线如图 2 所示,主要包括:采样点的波谱数据的获取、LightGBM 回归模型的构建、盐湖锂浓度反演三个步骤。

反演时需要用到盐湖实测数据和遥感影像数据。实测数据由中国地质科学院扎布耶盐湖野外观测站提供,采样时间为 2017 年 8 月 16 日。采样点根据《盐湖及盐类矿产地质勘察规范》的规定,使用 $2\text{ km}\times 2\text{ km}$ 控制网格,取样深度为 0.2 m。一次采样并化验后形成本次的实测数据,采样点位及锂浓度值如图 3 所示。本期实测数据为最新的实测数据,但是其中绝大多数是北湖数据,南湖只采集了一个点位的数据,因此以北湖为例进行方法研究。遥感影像数据采用中等分辨率的 Landsat-8 卫星影像—LC08_L1TP_143038_20170823_20170912_01_T1。整个实验过程在 ENVI、ArcGIS、JupyterNotebook 中实现,并通过平均相对误差(AVE)、均方根误差(RMSE)和卡方(R^2)进行模型精度评价。

2.1 采样点的波谱数据的获取

本文使用的地物波谱数据从遥感影像中获得,获取数据之前需要先对遥感影像进行辐射定标、FLAASH 大气校正和裁剪等一系列的预处理,纠正影像在获取中的辐射失真,使遥感影像更真实地反映地物,获取的波谱信息更加准确。由于 Landsat 数据波段较少,直接反应锂浓度的特征波段不明显,所以将遥感影像的所有波段均作为训练特征。预处理后,根据盐湖采样点的经纬度生成采样点矢量数据,将同时期遥感影像所有波段波谱值提取至采样点矢量属性中,形成模型训练数据(表 1)。

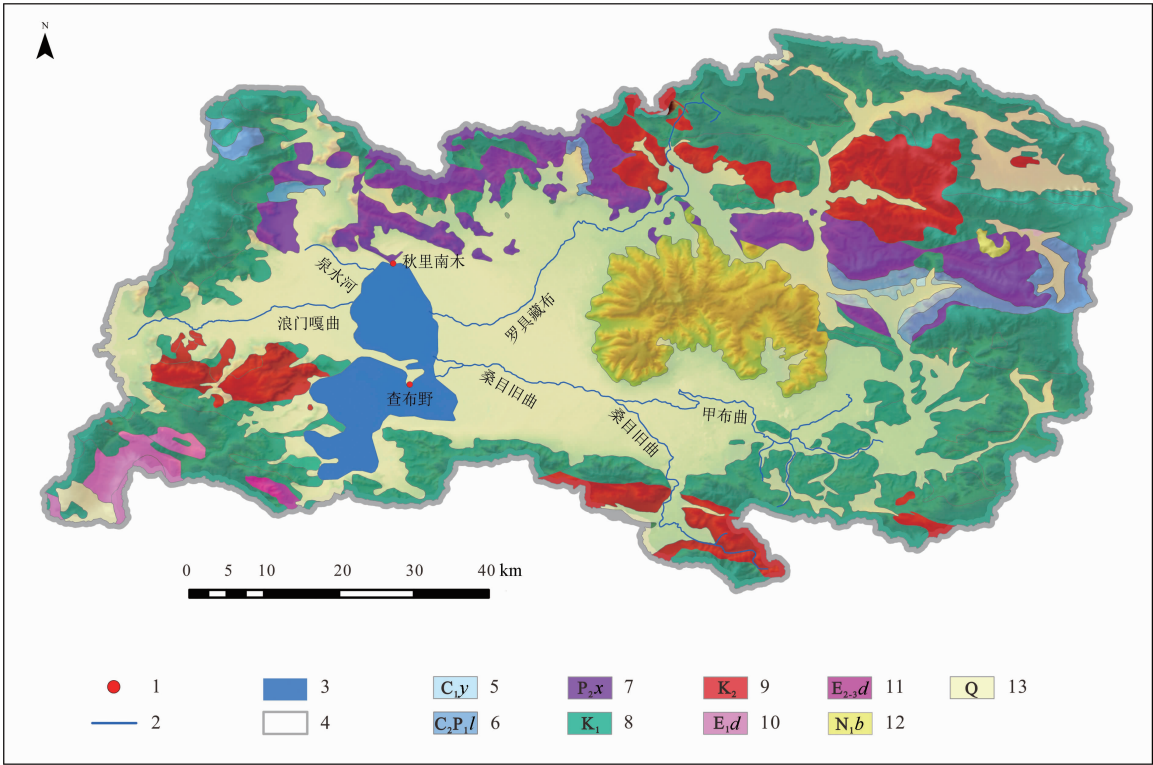


图 1 扎布耶流域地质略图

Fig. 1 Geological sketch of Zabuye basin

1—泉水;2—河流;3—湖泊;4—流域边界;5,6—石炭系(下石炭统,下二叠统一上石炭统);
7—二叠系(中二叠统);
8,9—白垩系;10,11—古近系(古新统,始新渐新统);12—新近系(中新统);13—第四系

1—Spring; 2—river; 3—lake; 4—basin boundary; 5,6—Carboniferous (Lower Carboniferous, Lower Permian—Upper Carboniferous);
7—Permian(Middle Permian); 8,9—Cretaceous; 10,11—Paleogene (Paleogene, Eocene Oligocene); 12—Neogene (Miocene); 13—Quaternary

2.2 LightGBM 回归模型的构建

本文采用 LightGBM 回归模型预测的不同波段特征下的锂浓度。LightGBM 是目前回归预测准确率较高的算法之一,内部集成了多棵决策树,能够综合多棵决策树的决策结果,从而避免单一学习器准确率不高的缺陷。除此之外,LightGBM 还采用直方图算法、带深度限制的叶子生长策略、GOSS 抽样方法、独特特征捆绑等方法提升整个算法的准确率和训练速度,并且能够自行筛选有效的数据特征(Zhang Ying et al. , 2019; Li Lijuan et al. , 2021; Song Jiazhi et al. , 2021),非常适合解决 Landsat 遥感数据波段较少带来的辐射值与锂浓度之间的映射关系复杂,全部波段作为训练数据带来的数据特征不明显等问题。

LightGBM 回归模型构建在 JupyterNotebook 中完成,构建过程如下:

(1)将盐湖锂浓度数据和波谱数据分为训练样本和验证样本。在采样点有限的情况下,为了保证训练样本能全面地反映整个湖面各种浓度锂的波谱特征,于是兼顾采样点的空间分布和锂浓度值高低

人为地进行了训练样本和验证样本的划分,将 27 个样本数据按照 9 : 1 的比例分为 24 个训练样本和 3 个验证样本。

(2)对 LightGBM 回归模型参数进行初始化。LightGBM 主要通过学习率(learning_rate)、训练轮数(num_boost_round)、树中最大的叶子结点数(num_leaves)、树的最大深度(max_depth)、最小分裂值(min_gain_to_split)、一个叶子上最小数据量(min_data_in_leaf)、迭代时的参数比例(feature_fraction)、迭代时的数据比例(bagging_fraction)、L1 正则化系数(reg_alpha)和 L2 正则化系数(reg_lambda)实现控制和优化的。本研究中 learning_rate 设置为 0.5,num_boost_round 设置为 50,num_leaves 设置为 20,max_depth 设置为 4,min_gain_to_split 设置为 0.01,min_data_in_leaf 设置为 16,feature_fraction 设置为 1,bagging_fraction 设置为 1,reg_alpha 设置为 0,reg_lambda 设置为 0。

(3)利用训练样本对 LightGBM 回归模型进行训练。具体训练过程包括初始梯度值的计算、决策树的建立以及梯度值的更新。决策树的建立又包括

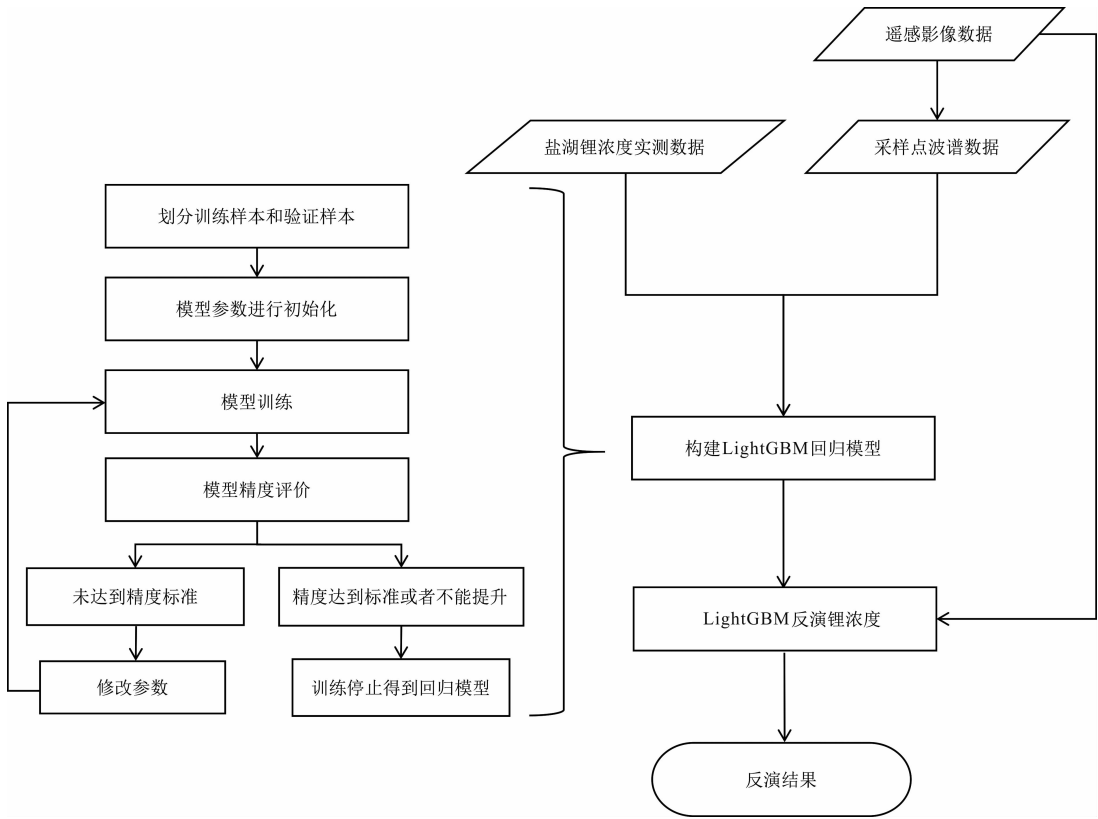


图 2 反演方法技术路线图

Fig. 2 Technology road map of inversion method

直方图的计算、最佳分裂特征的选取、根节点的建立、样本的切分、样本输出值的更新等过程(Zhou Wen et al., 2019)。

(4)利用验证样本对训练模型进行精度评价,若未达到精度标准则修改参数重新训练。依次修改 learning_rate 和 num_boost_round、num_leaves 和 max_depth、min_gain_to_split 和 min_data_in_leaf、feature_fraction 和 bagging_fraction 以及 reg_alpha 和 reg_lambda,从而找到最优的参数设置。

(5)反复训练直到精度达到标准或者不能提升,得到回归模型。本文最终的参数设置为:learning_rate 为 0.15,num_boost_round 为 90,num_leaves 为 20,max_depth 为 3,min_gain_to_split 为 0.01,min_data_in_leaf 为 10,feature_fraction 为 1,bagging_fraction 为 1,reg_alpha 为 0,reg_lambda 为 0。

2.3 盐湖锂浓度反演

锂浓度遥感反演是基于卤水波谱特性开展的,而提供卤水波谱特征的遥感影像最基本的构成单元是像元,因此要想利用回归模型预测整个盐湖的锂浓度值就得将遥感数据的每个像元变换成训练数据的形式输入模型中,即一个像元点所有波段的像元值作为一组波谱数据,遥感影像每个波段的像元值

均作为一个反演特征(图 4)。这一过程在 JupyterNotebook 中进行。首先读取遥感影像,进行遥感影像的格式变换和数据重新排列;然后送入训练好的 LightGBM 回归模型中,反演出每个像元点的锂浓度值;最后在将得到的锂浓度值按照之前波谱数据在遥感影像空间维上的位置分布进行反向的数据排列,并赋予坐标系、行列数、波段数等信息,输出 tiff 格式影像得到反演结果。

3 结果与分析

(1)LightGBM 回归模型精度评价结果(表 2)显示实测值与预测值的平均相对误差(AVE)为 0.053530925,均方根误差(RMSE)为 10.2869,卡方(R^2)为 0.867,模型与数据拟合程度较高,建立的回归模型有效。从反演结果来看(图 5),与实际情况吻合。整个北湖中,锂浓度最高的是东南部的水域,最低的是中西部的水域。桑目旧曲、罗具藏布、浪门嘎曲、泉水河等河流和秋里南木泉水群的汇入使得附近的水域锂浓度降低,但是对于本身位于中部锂浓度低值区的钙华岛来说,岛内泉水的汇入使得附近的水域锂浓度有所升高。

(2)截止目前,针对扎布耶盐湖进行的锂浓度遥

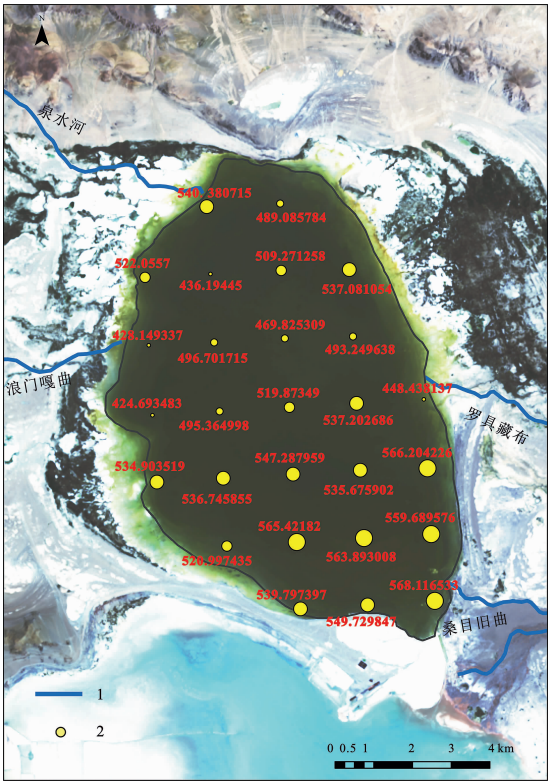


图 3 2017 年扎布耶盐湖采样点位分布图

Fig. 3 Distribution map of sampling sites in Zabuye salt lake in 2017

1—河流;2—采样点(锂浓度值:mg/L)

1—River;2—sample points (lithium concentration value: mg/L)

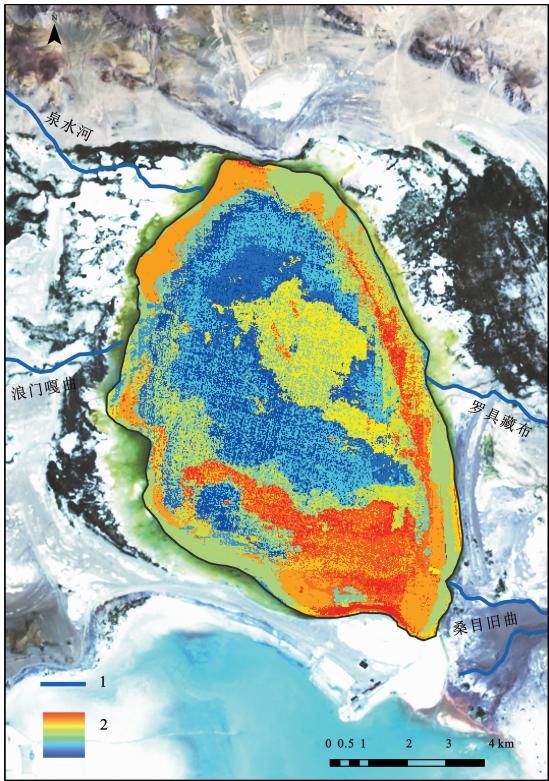


图 5 2017 扎布耶盐湖锂浓度反演结果图

Fig. 5 Results of lithium concentration inversion in Zabuye salt lake in 2017

1—河流;2—反演值(高:565.764 mg/L;低:466.01 mg/L)

1—River;2—value of inversion

(high:565.764 mg/L; low:466.01 mg/L)

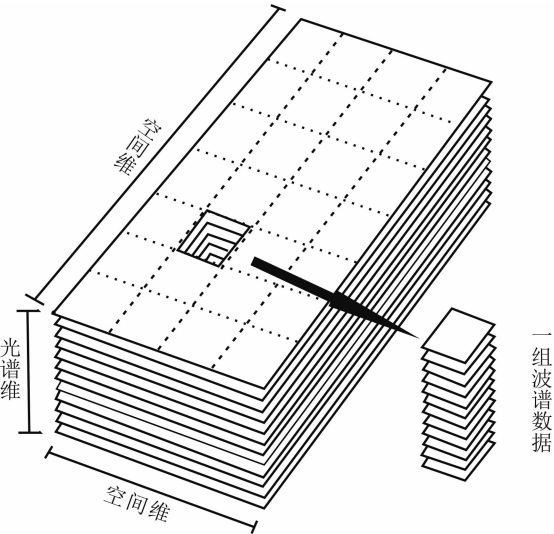


图 4 反演数据示意图

Fig. 4 Schematic diagram of inversion data

感反演使用的均为回归分析等数学统计方法,反演精度 R^2 最高为 0.773,校正后 R^2 为 0.762 (Tian Shufang et al. , 2005; Xu Weninng et al. , 2017),而本文使用的 LightGBM 机器学习算法反演精度

R^2 达到 0.867,校正后 R^2 为 0.815。在使用的遥感数据和采样数据基本一致的情况下,本文提供的 LightGBM 锂浓度反演方法精度更高。本研究提供的方法与现有的同类型的盐湖反演研究相比,也具有相对较高的反演精度(Wang Junhu et al. , 2015; Zhou Yamin et al. , 2016; Wang Dejun et al. , 2018; Wang Dejun, 2019)。

(3) 本文使用的遥感数据为中等分辨率的 Landsat-8 数据,波段较少,对反演结果也会有所影响。尽管使用的遥感数据空间和波谱分辨率均较低,但是模型预测精度较高。总的来看,LightGBM 用于盐湖锂浓度反演是可行的,且具有较高的精度。

(4) LightGBM 本身的优势使得前期的波谱分析可以省去,但是遥感反演需要实测数据作为支撑,针对没有实测数据的情况,还需要从波谱分析着手。实际地用短波红外、热红外波谱仪测量出不同浓度卤水的高光谱,并从中找到反映浓度的波长位置、波段深度、宽度、斜率、面积及对称度等信息,从而利用波谱直接预测浓度。也可以将青藏高原多个盐湖的

表 1 盐湖采样点波谱特征数据

Table 1 Spectral characteristic data of salt lake sampling points

编号	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	锂浓度值(mg/L)
1	318	174	932	858	−85	−21	−3	568.116533
2	185	57	722	607	−32	−13	−6	559.689576
3	257	119	807	645	−36	−7	−3	566.204226
4	295	189	1186	1239	−40	−7	−2	448.438137
5	259	145	682	535	1	−1	0	537.202686
6	289	165	731	579	4	−1	−3	493.249638
7	245	122	625	455	−27	−5	−1	537.081054
8	51	−44	611	523	−78	−7	1	549.729847
9	236	130	667	538	3	0	0	563.893008
10	236	125	664	537	1	−4	0	565.42182
11	394	236	819	722	−59	−15	0	539.797397
12	286	162	666	502	4	−1	−1	535.675902
13	291	157	642	473	5	0	2	547.287959
14	272	148	673	528	−2	−1	−2	519.87349
15	281	154	702	554	4	−7	−2	469.825309
16	305	171	689	530	2	−4	−1	509.271258
17	402	235	907	775	−29	−9	−4	520.997435
18	248	125	676	529	5	−6	−1	536.745855
19	274	151	662	514	3	−6	0	495.364998
20	286	157	683	529	3	−7	0	496.701715
21	313	175	662	497	2	−7	0	436.19445
22	224	88	779	632	−46	−16	−4	540.380715
23	139	30	888	876	−69	−22	−7	522.0557
24	270	142	688	518	−20	−11	−2	428.149337
25	237	110	633	476	−13	−9	0	424.693483
26	216	90	744	611	−19	−10	−2	534.903519
27	282	149	637	506	−23	−11	−4	489.085784

表 2 LightGBM 模型精度评价

Table 2 Model accuracy evaluation of LightGBM

FID	样品号	实际值(mg/L)	预测值(mg/L)	相对误差	备注
0	Z0101	568.116533	510.616	0.101212568	验证集
1	Z0201	559.689576	534.337	0.045297567	
2	Z0301	566.204226	546.615	0.034597456	
3	Z0401	448.438137	522.895	0.166035974	
4	Z0501	537.202686	525.68	0.02144942	
5	Z0601	493.249638	509.084	0.032102126	
6	Z0701	537.081054	534.969	0.003932468	
7	Z0801	549.729847	534.969	0.026851093	
8	Z0901	563.893008	551.954	0.02117247	验证集
9	Z1001	565.42182	551.954	0.023819067	
10	Z1101	539.797397	527.212	0.023315038	
11	Z1201	535.675902	497.438	0.071382532	
12	Z1301	547.287959	497.438	0.09108543	
13	Z1401	519.87349	480.842	0.075078824	
14	Z1501	469.825309	509.084	0.083560188	
15	Z1601	509.271258	494.885	0.028248714	验证集
16	Z1701	520.997435	510.616	0.019926077	
17	Z1801	536.745855	518.605	0.033797848	
18	Z1901	495.364998	497.438	0.004184797	
19	Z2001	496.701715	494.885	0.003657557	
20	Z2101	436.19445	497.438	0.140404239	
21	Z2201	540.380715	534.337	0.01118418	
22	Z2301	522.0557	534.337	0.023524884	
23	Z2401	428.149337	479.821	0.120686075	
24	Z2501	424.693483	508.879	0.198226534	
25	Z2601	534.903519	520.526	0.026878715	
26	Z2701	489.085784	482.374	0.013723122	
AVE : 0.053530925			RMSE : 10.2869		R ² : 0.867

实测数据形成大数据,利用机器学习算法形成模型,更好地预测青藏高原地区其他未建观测站的盐湖的矿物浓度信息,为盐湖观测节省大量的人力物力。

References

Fan Dexiang. 2018. Prediction and application of travel mode selection for urban residents using the Light Gradient Boosting Machine method. Master's thesis of Huazhong University of Science and Technology (in Chinese with English abstract).

Fan Jiazhi, Luo Yu, Tan Shiqi, Ma Wen, Zhang Honghao, Liu Fulai. 2021. Accuracy evaluation of the FY-3C/MWRI land surface temperature product in Hunan Province. *Remote Sensing of Land and Resources*, 33 (1): 249~255 (in Chinese with English abstract).

Kong Weigang, Bu Lingzhong, Zheng Mianping, Nie Zhen. 2017. Overview of salt lake stations on Tibetan Plateau. *Science & Technology Review*, 35 (6): 103~107 (in Chinese with English abstract).

Lei Li. 2019. Study on changes of lake areas since 2000 and it response to the climatic factors in Tibetan Plateau. Master's thesis of Chengdu University of Technology (in Chinese with English abstract).

Li Lijuan, Li Yiken, Yu Dexin, Liu Zhu Yun, Gao Yanjing, Qian Jianping. 2021. A multi-organ fusion and LightGBM based radiomics algorithm for high-risk esophageal varices prediction in Cirrhotic Patients. *IEEE Access*, 9: 15042~15052.

Liu Fan, Wang Chuankuan, Wang Xingchang. 2018. Application of near-surface remote sensing in monitoring the dynamics of forest canopy phenology. *Chinese Journal of Applied Ecology*, (6): 1768~1778 (in Chinese with English abstract).

Liu Xifang, Zheng Mianping. 1999. Geological setting and mineralizing evolution of super large Li, B deposit, Zabuye, Tibet. *Geology of Chemical Minerals*, 21 (2): 65~68 (in Chinese with English abstract).

Liu Xifang, Zheng Mianping, Qi Wen. 2007. Sources of ore-forming materials of the superlarge B and Li deposit in Zabuye Salt Lake, Tibet, China. *Acta Geologica Sinica*, 81 (12): 1709~1715 (in Chinese with English abstract).

Lu Shanlong, Xiao Gaohuai, Jia Li, Zhang Wei, Luo Haijing. 2016. Extraction of the spatial-temporal lake water surface dataset in the Tibetan Plateau over the past 10 years. *Remote Sensing of Land and Resources*, 28 (3): 181~187 (in Chinese with English abstract).

Pan Xuepeng, Li Gaixin, Liu Fenggui, Wu Xifang, Kondoh Akihiko, Shen Yanjun. 2015. Using remote sensing to determine spatio-temporal variations in winter wheat growing area in the North China Plain. *Chinese Journal of Eco-Agriculture*, 23 (4): 497~505 (in Chinese with English abstract).

Pan Zhenbao, Fang Shuhua, Wang Hui. 2021. LightGBM technique and differential evolution algorithm-based multi-objective optimization design of DS-APMM. *IEEE Transactions on Energy Conversion*, 36 (1): 441~455.

Qi Wen, Zheng Mianping. 2006a. Winters and ARIMA model analysis of the lake level of Salt Lake Zabuye, Tibetan Plateau. *Journal of Lake Sciences*, 18 (1): 21~28 (in Chinese with English abstract).

Qi Wen, Zheng Mianping. 2006b. Time-serial analyses of water level fluctuation of Zabuye Salt Lake, Tibet. *Scientia Geographica Sinica*, 26 (6): 693~700 (in Chinese with English abstract).

Song Jiazhi, Liu Guixia, Jiang Jingqing, Zhang Ping, Liang Yuanchun. 2021. Prediction of Protein-ATP binding residues based on ensemble of Deep Convolutional Neural Networks and LightGBM algorithm. *International Journal of Molecular Sciences*, 22 (2): 939.

Sun Yue. 2019. Chinese text classification based on LightGBM.

Master's thesis of Soochow University (in Chinese with English abstract).

Tian Shufang, Qin Xuwen, Zheng Mianping, Hong Youtang, Kuang Shengai. 2005. Quantitative analysis of Remote Sensing on the total salinity of Zabuye Salt Lake in Tibet. *Geoscience*, 19 (4): 506~602. (in Chinese with English abstract).

Wang Dejun, Jiang Qigang, Xi Jing, Lv Danhong. 2018. Quantitative analysis of lithium resources in Arizaro salt lake using remote sensing based on Sentinel-2A. *Global Geology*, 37 (4): 1250~1257 (in Chinese with English abstract).

Wang Dejun. 2019. Models for predicting the Li content in salt lake based on remote sensing: a case study of Argentina's Arzaro Salt Lake. Master thesis of Jilin University (in Chinese with English abstract).

Wang Junhu, Liu Jia, Li Zhizhong, Liu Dechang, Wang Daming. 2015. High resolution remote sensing estimation of salinity in salt lake with uranium resoources. *Earth Science-Journal of China University of Geosciences*, 40 (8): 1409~1414 (in Chinese with English abstract).

Xie Yong, Xiang Wei, Ji Mengzhong, Peng Jun, Huang Yihuai. 2019. An application and analysis of forecast housing rental based on Xgboost and LightGBM algorithms. *Computer Application and Software*, 36 (9): 152~155, 191 (in Chinese with English abstract).

Xu Wenning, Bu Lingzhong, Kong Weigang, Zheng Mianping, Nie Zhen. 2017. Monitoring of the dynamic change of Zabuye Salt Lake: a remote sensing approach. *Science & Technology Review*, 35 (6): 89~96 (in Chinese with English abstract).

Yang Kehan, Yao Fangfang, Dong Di, Dong Wen, Luo Jiancheng. 2017. Spatiotemporal monitoring of lake area dynamics on the Tibetan Plateau. *Journal of Geo-information Science*, 19 (7): 972~982 (in Chinese with English abstract).

Zhang Chungui, He Jinde, Ma Zhiguo. 2013. Remote sensing monitor of sea fog in Fujian coastal region. *Chinese Journal of Agrometeorology*, 34 (3): 366~373 (in Chinese with English abstract).

Zhang Ying, Wang Yanhao, Gao Minghe, Ma Qunfei, Zhao Jing, Zhang Rongrong, Wang Qingqing, Huang Linyan. 2019. A predictive data feature exploration-based air quality prediction approach. *IEEE Access*, 7: 30732~30743.

Zhou Wen, Wang Yu, Li Changsheng, Xiao Hongbin, Xing Suxia. 2019. Application of LightGBM algorithm in classification of patients with Alzheimer's disease from structural magnetic resonance images. *Chinese Journal of Medical Physics*, 36 (4): 408~413 (in Chinese with English abstract).

Zhou Yamin, Zhang Rongqun, Ma Hongyuan, Zhang Jian, Zhang Xiaoshuan. 2016. Retrieving of salt lake mineral ions salinity from hyper-spectral data based on BP neural network. *Remote Sensing of Land and Resources*, 28 (2): 34~40 (in Chinese with English abstract).

参 考 文 献

范德祥, 2018. 基于 LightGBM 的居民出行方式选择模型及应用研究. 华中科技大学硕士学位论文.

范嘉智, 罗宇, 谭诗琪, 马雯, 张弘豪, 刘富来. 2021. 基于 FY-3C/MWRI 的湖南省地表温度遥感反演评价 国土资源遥感, 33 (1): 249~255.

孔维刚, 卜令忠, 郑绵平, 乜贞. 2017. 青藏高原盐湖观测站概述. 科技导报, 35 (6): 103~107.

闰利. 2019. 2000 年以来青藏高原湖泊面积变化与气候要素的影响关系研究. 成都理工大学硕士学位论文.

刘帆, 王传宽, 王兴昌. 2018. 近地遥感在森林冠层物候动态监测中的应用. 应用生态学报, (6): 1768~1778.

刘喜方, 郑绵平. 1999. 西藏扎布耶超大型锂硼矿床地质背景及其形成演化. 化工矿产地质, 21 (2): 65~68.

刘喜方, 郑绵平, 齐文. 2007. 西藏扎布耶盐湖超大型 B、Li 矿床成矿物质来源研究. 地质学报, 81 (12): 1709~1715.

卢善龙, 肖高怀, 贾立, 张微, 罗海静. 2016. 2000-2012年青藏高原湖泊水面时空过程数据集遥感提取. 国土资源遥感, 28(3): 181~187.

潘学鹏, 李改欣, 刘峰贵, 吴喜芳, 近藤昭彦, 沈彦俊. 2015. 华北平原冬小麦面积遥感提取及时空变化研究. 中国生态农业学报, 23(4): 497~505.

齐文, 郑绵平. 2006a. 西藏扎布耶盐湖水位 Winters 和 ARIMA 模型分析. 湖泊科学, 18(1): 21~28.

齐文, 郑绵平. 2006b. 西藏扎布耶盐湖水位波动规律初探. 地理科学, 26(6): 693~700.

田淑芳, 秦绪文, 郑绵平, 洪友堂, 邝生爱. 2005. 西藏扎布耶盐湖总盐含量遥感定量分析. 现代地质, 19(4): 596~602.

孙悦. 2019. 基于 LightGBM 的中文文本分类研究. 苏州大学硕士学位论文.

王俊虎, 刘佳, 李志忠, 刘德长, 汪大明. 2015. 含铀盐湖矿化度高分辨率遥感估测. 地球科学——中国地质大学学报, 40(8): 1409~1414.

王德军, 姜琦刚, 习靖, 吕丹红. 2018. 基于 Sentinel-2A 的阿里萨

罗盐湖锂资源遥感定量分析. 世界地质, 37(4): 1250~1257.

王德军. 2019. 盐湖 Li 含量遥感预测模型研究——以阿根廷阿里萨罗盐湖为例. 吉林大学硕士学位论文.

谢勇, 项薇, 季孟忠, 彭俊, 黄益槐. 2019. 基于 Xgboost 和 LightGBM 算法预测住房月租金的应用分析. 计算机应用与软件, 36(9): 152~155, 191.

许文宁, 卜令忠, 孔维刚, 郑绵平, 乜贞. 2017. 扎布耶盐湖动态变化遥感监测. 科技导报, 35(6): 89~96.

杨珂含, 姚方方, 董迪, 董文, 骆剑承. 2017. 青藏高原湖泊面积动态监测. 地球信息科学, 19(7): 972~982.

张春桂, 何金德, 马治国. 2013. 福建沿海海雾的卫星遥感监测. 中国农业气象, 34(3): 366~373.

周文, 王瑜, 李长胜, 肖洪兵, 邢素霞. 2019. LightGBM 算法在阿尔茨海默症结构磁共振成像分类中的应用. 中国医学物理学杂志, 36(4): 408~413.

周亚敏, 张荣群, 马鸿元, 张健, 张小栓. 2016. 基于 BP 神经网络的盐湖矿物离子含量高光谱反演. 国土资源遥感, 28(2): 34~40.

Remote sensing inversion of lithium concentration in salt lake using LightGBM: a case study of northern Zabuye salt lake in Tibet

LIU Tingyue¹⁾, DAI Jingjing^{*2)}, ZHAO Yuanyi²⁾, TIAN Shufang¹⁾, YE Chuanyong²⁾

1) School of Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China;

2) Institute of Mineral Resources, Chinese Academy of Geological Sciences, Beijing 100037, China

* Corresponding author: daijingjing863@sina.com

Abstract

The mineral resources such as lithium and boron in the saline lakes of Qinghai Tibet Plateau are of great economic value. The content and spatial distribution of lithium, boron and other mineral resources are the focus of salt lake observation, which has important significance for the development and utilization of saline lake resources. Using remote sensing technology to carry out saline lake observation can overcome the shortcomings of traditional observation station, such as discontinuous data, time and labor intensive, while machine learning algorithms can quickly and efficiently mine remote sensing data information. Therefore, based on landsat-8 remote sensing image data, this paper uses LightGBM algorithm to carry out remote sensing inversion of lithium concentration in north lake of Zabuye salt lake in Tibet. After a series of experimental steps, such as the acquisition of spectral data of sampling points, the construction of LightGBM regression model of lithium concentration, and the inversion of lithium concentration in salt lake, the final model evaluation results show that the average relative error is 0.053530925, the root mean square error is 10.2869, and the chi square is 0.867. The model has a high degree of fitting with the measured data. The inversion results show that: in the whole of north lake the highest lithium concentration is in the southeast waters, and the lowest is in the midwest waters; the confluence of rivers and the Qiulinanmu spring group lowers the lithium concentration in the nearby waters; for travertine island, which is located in the low lithium concentration area in the middle, the confluence of springs in the island makes the lithium concentration in the nearby waters higher, which is consistent with the actual situation. Through the inversion research in this paper, the feasibility and accuracy of LightGBM machine learning algorithm for fast inversion of saline lake lithium concentration are proved. At the same time, it provides technical explanation for remote sensing inversion of other salt lake mineral resources, and lays the foundation for subsequent salt lake resource assessment.

Key words: Qinghai-Tibet Plateau; Zabuye salt lake; lithium concentration; remote sensing inversion; LightGBM