

基于循环神经网络的找矿模型构建与预测

张亚光^{1,2}, 陈建平^{1,2}, 贾志杰^{1,2}, 李 诗^{1,2}, 刘苏庆^{1,2}, 张志平³, 张 焯³
ZHANG Yaguang^{1,2}, CHEN Jianping^{1,2}, JIA Zhijie^{1,2}, LI Shi^{1,2}, LIU Suqing^{1,2},
ZHANG Zhiping³, ZHANG Ye³

1. 中国地质大学(北京)地球科学与资源学院, 北京 100083;

2. 北京市国土资源信息研究开发重点实验室, 北京 100083;

3. 页岩气勘探开发国家地方联合工程研究中心(重庆地质矿产研究院), 重庆 401120

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

2. Beijing Key Laboratory of Development and Research for Land Resources Information, Beijing 100083, China;

3. National and Local Joint Engineering Research Center of Shale Gas Exploration and Development, Chongqing Geology and Mineral Research Institute, Chongqing 401120, China

摘要:在大数据和人工智能背景下,基于已有的传统地质找矿模型建立与应用基础,提出基于循环神经网络的找矿模型构建与预测方法,实现对地质数据的深入分析和理解。针对地质找矿模型构建与预测的需求,结合数据清洗理论,对传统地质找矿模型进行归纳与总结,建立地质找矿知识库,为深度学习算法提供训练数据。通过分类算法研究,综合对比结果的准确率与分类所用时间,最终选用RNN分类算法对找矿概念模型进行分类。在建立研究区找矿模型中,通过关键词与控矿要素完成模型匹配,利用模型计算对模型匹配结果进行数据分析,实现区域地质找矿模型的构建与矿产资源的预测评价和分析。以大水金矿为例,快速准确地实现了找矿模型的构建,有效地对矿产资源预测工作提供了指导,验证了该方法的可行性。

关键词:找矿概念模型;RNN分类算法;控矿要素;模型匹配

中图分类号:P628 文献标志码:A 文章编号:1671-2552(2019)12-2033-10

Zhang Y G, Chen J P, Jia Z J, Li S, Liu S Q, Zhang Z P, Zhang Y. Construction and prediction of a prospecting model based on recurrent neural network. *Geological Bulletin of China*, 2019, 38(12): 2033-2042

Abstract: Under the background of big data and artificial intelligence and on the basis of the establishment and application basis of existing traditional geological prospecting model, this paper proposes a prospecting model construction and prediction method based on cyclic neural network, with the purpose of achieving in-depth analysis and understanding of geological data. According to the requirements for construction and prediction of geological prospecting model, the authors combined the data cleaning theory to systematically summarize and summarize the traditional geological prospecting model, thus establishing a geological prospecting knowledge base and providing training data for deep learning algorithms. The accuracy of the comparison results and the time used for classification were comprehensively analyzed. Finally, the RNN classification algorithm was selected to classify the conceptual model of prospecting. In the process of establishing the prospecting model of the study area, by using the key words and ore control elements to complete the model matching, the model was used to analyze the model matching results so as to realize the construction of the regional geological prospecting model and the prediction and analysis of the mineral resources. With the Dashui gold deposit as an example, the construction of the prospecting model was realized quickly and accurately, which effectively provides guidance for the prediction of mineral resources and verifies the feasibility of the method.

收稿日期:2019-04-17;修订日期:2019-07-25

资助项目:国家科技部深地资源勘探开采专项《深部成矿地质异常定量预测方法与模型》(编号:2017YFC0601502)

作者简介:张亚光(1995-),男,硕士,从事地质大数据与矿产资源评价研究。E-mail: 410778030@qq.com

通讯作者:陈建平(1959-),男,教授,博士生导师,从事矿产资源定量评价研究工作。E-mail: 3s@cugb.edu.cn

Key words: prospecting model; RNN classification algorithm; ore control elements; model matching

在中国现代化建设中,矿产资源占据十分重要的战略地位。随着社会的发展,中国对矿产资源的需求量持续扩大,但现有矿产资源逐渐减少,找矿难度也不断加大。必须对矿产资源进行预测评价,实现指导国家政府资源政策制定和矿产勘查部署工作^[1]。面对如此严峻的挑战,不但需要以科学的资源观、求异的思维、系统的观念和超前的意识为指导,而且必须重视新理论、新技术、新方法的联合攻关^[2]。随着大数据呈现出势不可挡地发展趋势,大数据已得到世界范围的广泛关注,并渗透到生活的各个方面^[3-4]。在大数据背景下,不论是地理信息领域还是地质领域都有了新的驱动力^[5]。在地质领域,随着地质信息化的不断深入,地质数据的数据量飞速增长。地质数据作为地质工作形成的重要基础信息资源^[6],其范围涉及地球科学的各个方面,例如地球形成与演化的历史、地球的物质组成及其变化、矿产资源的形成、勘查与开发利用、人类环境的破坏与修复等。

地质大数据是信息时代下大数据的理念、技术和方法在地质领域的应用与实践。地质大数据的存在形式是地质数据之间复杂的网络关系,只有深入分析地质大数据关系网络,才能把握地质大数据的本质^[7]。但其深度价值通过简单分析很难发现,因此通常需要使用基于机器学习和数据挖掘的智能化复杂分析才能实现^[8]。同时机器学习的过程是一个极复杂的认知过程,其研究主旨是使用计算机模拟人类的学习活动^[9-10]。而作为机器学习中更深层次的深度学习,则是通过利用深度神经网络解决特征表达的一种学习过程,通过对激活函数或神经元的链接方式做出部分调整,提高深层神经网络的训练效果。在深度学习的基础上,结合地质科学时空大数据,应用地质科学、大数据、深度学习等技术,提出找矿模型构建与预测新方法,将找矿难度降低。通过对研究区成矿规律与成矿有关的各种空间信息进行综合分析、解释与总结,最终实现找矿靶区圈定和资源量估算^[11-13]。

1 找矿模型预测理论及技术方法

找矿模型在成矿预测中占据主导地位^[14]。不论其数据或资料完善与否,都是以经验模型和理论模型的综合信息融合为基础,以找矿为目的,以各类基础资料为基本内容,以不同控矿要素为依据,形

成找矿模型预测方法指导找矿。从相关的文献资料中可以分析主要矿床类型、找矿标志及控矿要素^[15]。而基于深度学习的找矿模型预测方法以矿床成矿理论作为理论依据。目前主要的矿床成矿理论包括翟裕生等的成矿系统理论^[16]、涂光炽等的大型与超大型矿床的成矿和找矿理论^[17]、朱裕生等的矿床成矿模式理论^[18]、程裕淇等的矿床成矿系列理论^[19]等。通过搜集国内外各类找矿模型,建立统一且容易区分的找矿地质概念模型库,以各研究区的勘查数据资料为基础;通过深度学习将找矿概念模型库中的找矿模型及控矿要素与研究区数据资料进行分析和归纳,构建找矿预测模型;将找矿预测模型中控矿要素确定后,依据研究区范围提供数据资料整理清单,完善找矿概念模型的数据基础;依据立方体定量预测系统中总结出的算法,推荐控矿要素适合的算法组合,最终在找矿概念预测模型的基础上,实现定量、定位和定概率的预测评价。其技术方法如图1所示。

2 找矿模型预测方法

2.1 找矿知识库构建

从大多已知矿区的找矿模型看,预测区的选择通常取决于该成矿区带的大地构造背景。大多已知矿床的找矿标志、成矿条件及控矿要素对找矿模型的构建起决定性作用,且大多已知矿床的形成时代也决定着成矿期的确定。因此,找矿知识库的构建主要包含3部分:找矿概念模型库的构建、工作程度模型库的构建及已知矿产地模型库的构建。

2.1.1 找矿概念模型库

找矿概念模型库作为找矿知识库的一个重要组成部分,它的构建为深度学习分类提供了重要的数据基础。在大数据深度学习中,构建找矿概念模型库时,需要建立易于理解、结构统一的属性数据库及适用于所有矿床成矿模式的找矿概念模型数据结构,为找矿概念模型的建立提供训练数据、相关参考文献、模型名称参考等。

通过对数据资料进行研究分析发现,在建立找矿模型时,由于资料来源不同及成矿地质条件和矿产勘查程度存在差异,导致对其理解和具体操作出现不统一的结果,例如,不同概念由同一名称表示,同一内涵由不同名称表达。因此,必须对所有控矿

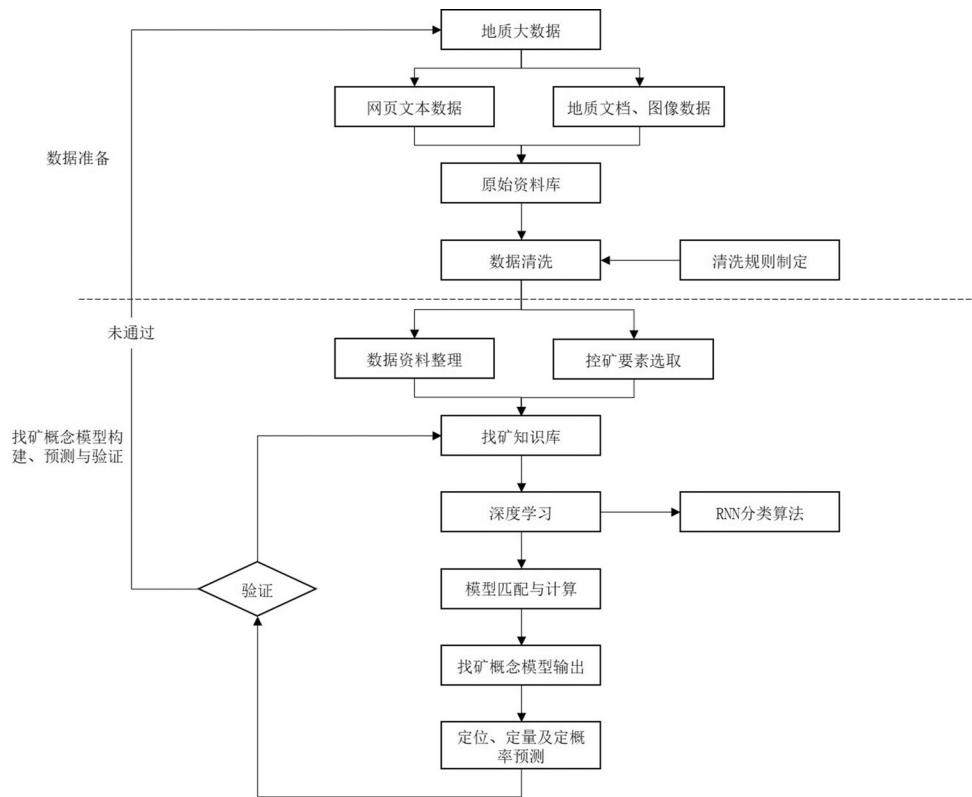


图 1 基于深度学习找矿模型构建与预测技术方法

Fig. 1 Construction and prediction of prospecting model based on in-depth learning

要素进行统一整理,建立找矿模型与控矿要素多对多的关系(图2)。

在对模型数据进行整理时,需要对模型名称与控矿要素进行整理。模型名称分为典型式命名与抽象式命名2种类型,在数据整理阶段二者无法统一,只能将其中的关键词与非关键词尽量完成统一。随着收集的模型越来越多,同一控矿要素会重复使用,针对这一情况,必须保证同一控矿要素的唯一性。

2.1.2 工作程度数据模型库

工作程度数据模型库的构建需建立工作区的空间模型,使其能结合空间概念理解和分析找矿概念模型,并有效利用对应的空间数据资料。研究区的空间模型需要存储大量的文件数据,可采用 GridFS 方式将大型文件存储在 MongoDB 中。其原理为将一个大文件分割为多个可存储的较小文档,并通过复制保障数据故障、数据集成等,通过分片实现数据的自动切分,保证大数据的存储和负载均衡。

2.1.3 已知矿产地模型库

在找矿概念模型库的基础上,添加已知矿产地坐标,构建已知矿产地模型库,同时结合属性模型和空间模型,在专家系统的基础上,确定矿床形成的成矿地质背景,为找矿概念模型构建提供更全面及有效的数据信息支撑,实现找矿模型的特征对比和统计分析。

2.2 找矿概念模型分类

找矿概念模型以文本数据为基础,通过分类处理的方法对其进行预测。基于深度学习的地质找矿模型分类方法是在已有找矿概念模型的基础上,利用地质文本展开分类工作。通过文本分类确定研究区的找矿模型属于哪个评价模型,并利用控矿要素的重要性与使用率判定相似典型矿床模型的排序。针对地质文本数据的分类算法有多种^[20],例如,Rocchio 算法、朴素贝叶斯、CNN 算法、RNN 算法等。在实验中,利用不同的分类算法对收集整理得到的 100 个数据进行分类,结果如图 3 和图 4 所示。从图 4 可以看出,RNN 进行分类的时间最短、

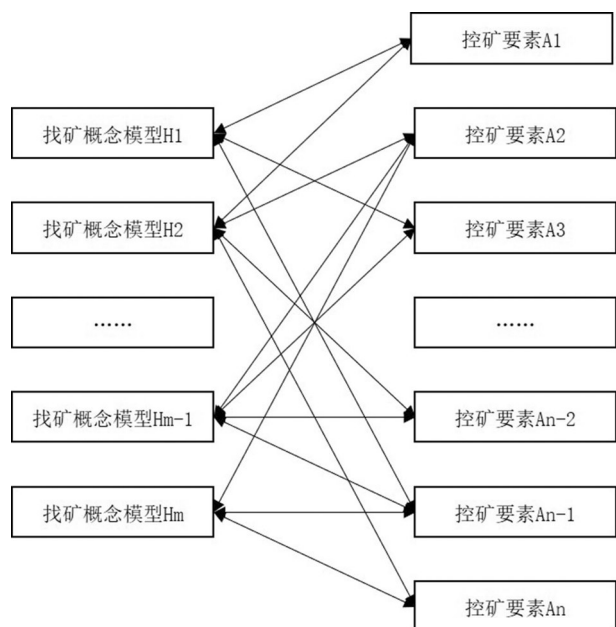


图2 找矿概念模型数据库中对应关系

Fig. 2 Correspondence relations in conceptual model database of prospecting

效率最高,虽然其准确率不如CNN算法(图3),但是通过对分类结果的准确率与分类所用时长进行综合分析,更加倾向于RNN分类算法。因此,本文主要选用RNN算法对地质文本数据进行分类。

2.2.1 朴素贝叶斯分类

朴素贝叶斯文本分类方法的训练样本为找矿概念模型库中的现有数据,将经过整理的研究区控矿要素作为待处理的数据,对研究区资料进行分类,计算对研究区控矿要素的条件概率,判断其属于模型库中每个模型的概率^[21]。朴素贝叶斯具有稳定的分类效率,算法较简单,非常适用于进行文本分类,尤其对小规模数据分类效果更好,且还能将其用于处理多分类任务。但该算法是通过先验数据、后验概率而决定分类的,导致分类决策具有一定的错误率。

2.2.2 CNN算法分类

CNN算法分类主要用来处理报告、文献等地质文本数据。在已有的找矿概念模型库中现有数据的基础上,利用词向量构建二维文本数据,将其作为训练样本输入,实现对找矿概念模型特征的提取与分类。主要针对原始文本数据,获取文本之间的相关关系,实现找矿模型的构建,并对其进行分

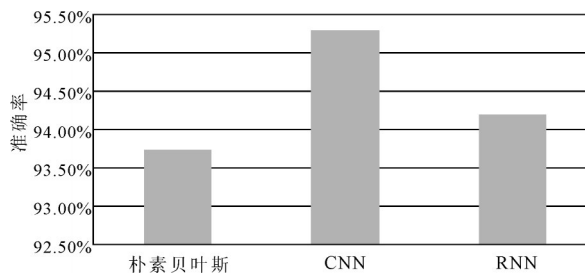


图3 分类结果准确率对比

Fig. 3 Comparison of accuracies of classification results

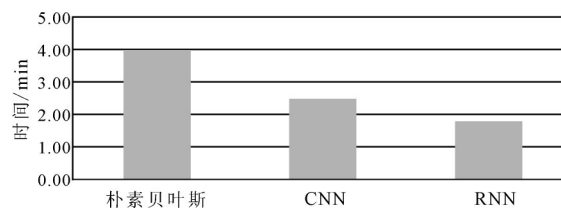


图4 分类所用时间对比

Fig. 4 Time comparison for classification

析与预测。用CNN算法来处理地质文本数据可以极大地避免数据预处理工作,还具有较强的特征提取能力^[22]。但从以往的深度学习算法应用看,CNN算法主要用于处理图像数据,其训练的样本量大,参数需要经过多次调整。

2.2.3 RNN算法分类

RNN是一类用于处理序列数据的神经网络^[23]。与基础的神经网络相比,RNN不仅在层与层之间建立了权连接,在层之间的神经元之间也建立了权连接。在RNN算法中,权值是共享的,且每一个输入值都只与其本身的那条路线建立权连接,不与别的神经元连接^[24-26]。此外,循环神经网络做地质文本分类任务,能够提取到词与词之间的关系^[27-28]。

地质文本数据具有文字序列,在利用RNN算法进行分类之前,需对地质文本数据进行预处理。预处理的目的是将文本转化为适合于表示和分类的干净的词序列^[29-30]。其过程一般包括以下几个步骤:①对地质文本数据进行清洗,减少不必要的计算;②确保找矿概念模型的一致性,如将所有的字母都用大写、删除地质文本数据中的空格与某些字符;③确保有用性,提取地质文本中的实词,去除各类虚词副词等;④保证文本数据的序列性,对删改后的地质文本进行重新排序,确保前后文本数据存

在一定的关系。利用预处理后的已有典型矿床模型与预测评价模型相关的文本,通过RNN深度学习训练,实现自动分析地质文本数据,并提取其文本抽象特征。

RNN中一些重要的设计模式包括以下几种:
 ①每执行一个时间步都有输出,在隐藏单元内部具有循环连接的一种循环网络。如图5所示。损失 L 是用来衡量每个 o 与对应训练目标 y 的距离。当利用softmax进行输出时,假设 o 是未归一化的对数概率,计算 L 内部计算 $\hat{y} = \text{soft max}(o)$,并将其与目标 y 比较。
 ②每个时间步都产生一个输出,当前时刻的输出到下个时刻的隐藏单元之间有循环连接的循环网络,它的唯一循环是从输出到隐藏层的反馈连接。在这种循环网络方式中,会导致 o 缺乏过去的重要信息,这也使RNN的功能不再那么强大,但更加容易对其进行训练,训练期间可以出现更多的并行化。
 ③各隐藏单元之间存在循环连接,但需读取整个序列后才能产生单个输出的循环网络,如图6所示。该类网络可以用于概括序列,还可用于处理固定大小表示,当出现与目标差距较大时,不利于检验其错误。

图6中, x 是一个向量,表示输入层的值; h 是一个向量,表示隐藏层的值,该层中存在多个节点,节

点数与向量 h 的维度相同; U 是输入层到隐藏层的权重矩阵; o 也是一个向量,表示输出层的值; V 是隐藏层到输出层的权重矩阵;循环神经网络的隐藏层的值 h 不仅取决于当前这次的输入 x ,还取决于上一次隐藏层的值 h ,权重矩阵 W 就是隐藏层上一次的值作为这一次的输入的权重。当 t 时刻接收到输入 x 之后,隐藏层的值是 h ,输出值是 o 。关键一点是, h 的值不仅取决于 x 值,还取决于 $t-1$ 时刻的 h 值,如图7所示,该图能更好地体现RNN的序列性。

RNN向前传播算法:从特定的初始状态 $h^{(0)}$ 开始向前传播,选定某段预处理后的地质文本数据,该数据由 n 个词组成,形成一个序列数据,输入时间步,从 $t=1$ 到 $t=\tau$ 的每个时间步运用以下公式进行方程的更新,达到训练分类的效果。

t 时刻隐藏单元值:

$$h^{(t)} = \varphi(Ux^{(t)} + Wh^{(t-1)} + b) \tag{1}$$

其中, $\varphi()$ 为激活函数,一般来说会选择tanh函数, b 为偏置。

t 时刻的输出值:

$$o^{(t)} = Vh^{(t)} + c \tag{2}$$

最终模型的预测输出值:

$$\hat{y}^{(t)} = \sigma(o^{(t)}) \tag{3}$$

其中, $\sigma()$ 为激活函数,通常RNN用于分类,

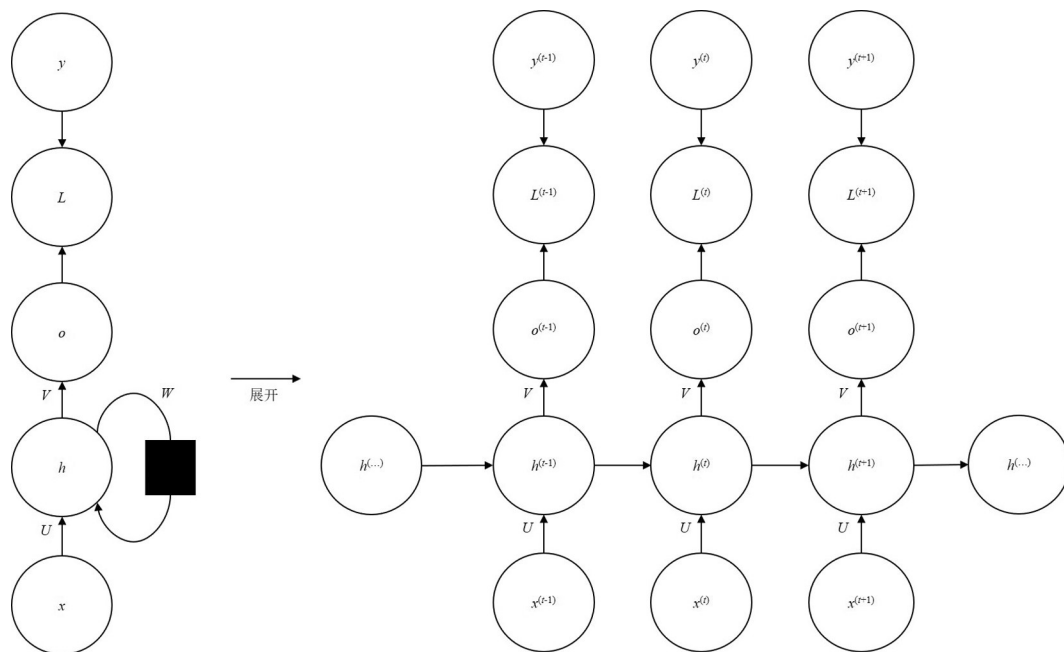


图5 循环神经网络训练损失计算图

Fig. 5 Calculating chart of training loss of recurrent neural network

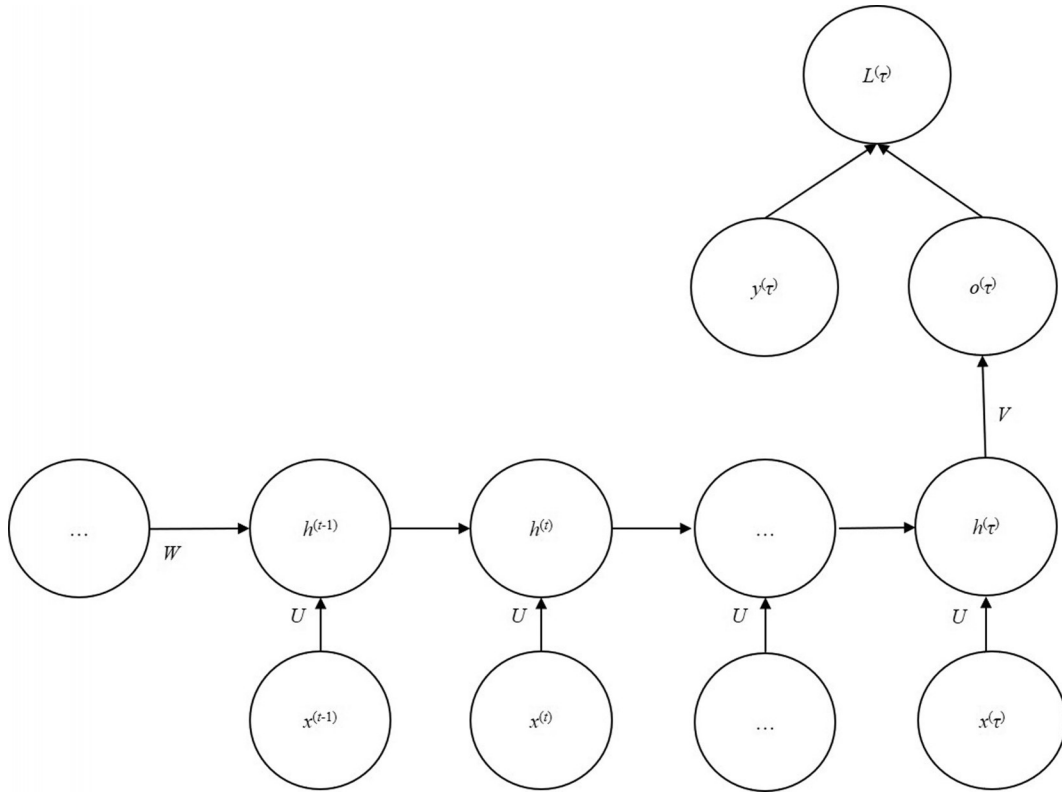


图6 关于时间展开的循环神经网络

Fig. 6 Recurrent neural network on time expansion

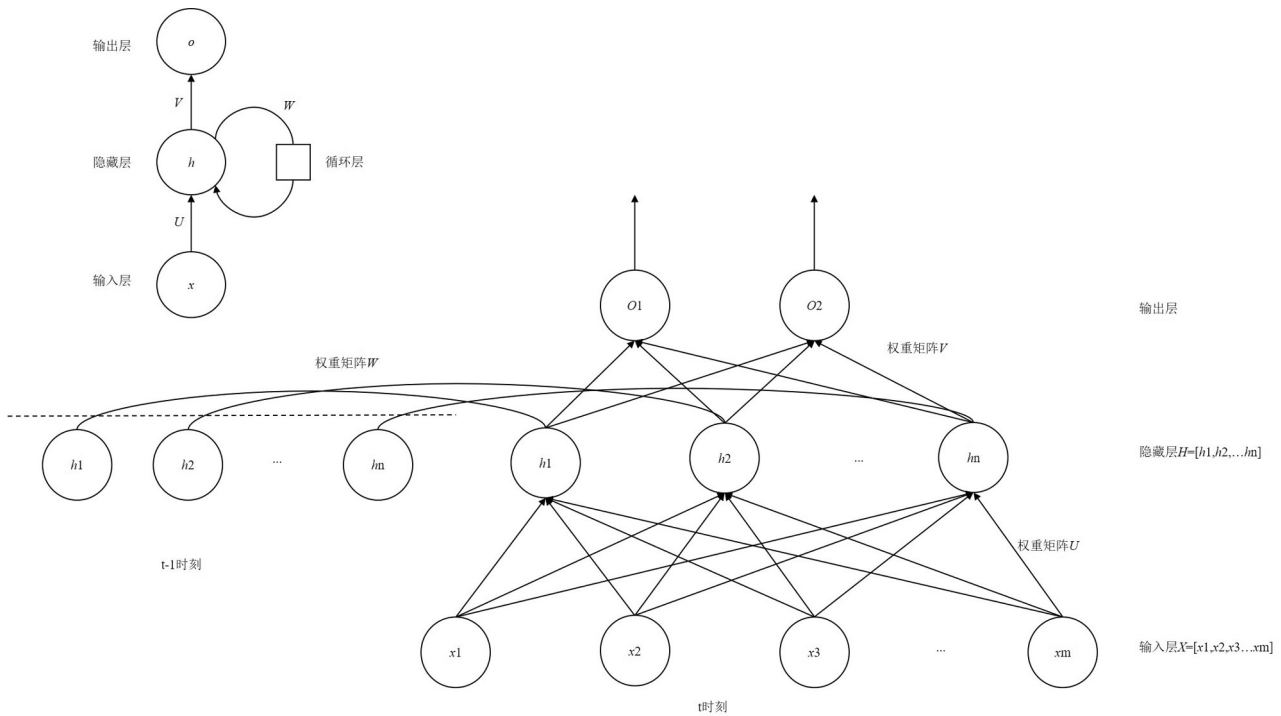


图7 RNN算法总括图

Fig. 7 Summary diagram of recurrent neural network

故这里一般用 softmax 函数。

在 RNN 文本分类算法中,将准备好的找矿概念模型数据作为训练样本,对研究区的控矿要素数据进行整理,并作为待处理的数据。从找矿概念模型库中找出相关联的模型,计算其控矿要素的重要性与使用率对库中的模型进行分级。综合相关联的模型数据构建找矿模型,实现对找矿模型中矿产资源的预测。

RNN 算法进行分类时,用的训练方法是 BPTT,即随时间反向传播,其本质是 BP 算法,沿着需要优化的参数的负梯度方向不断寻找更优的点直至收敛,RNN 训练需要优化的参数为 V 、 U 、 W 。分别对 V 、 U 、 W 三个权重矩阵求偏导,偏导求解过程中会嵌套激活函数,且 RNN 的损失也会随着时间累加,所以不能只求 t 时刻的偏导。

2.3 模型匹配与计算

对某一研究区进行资料收集、预处理后,分析研究区的情况,剥离出地质文本数据中关键词与控矿要素。关键词的选取主要根据地质词典的分词结果人工获取,利用多个关键词与训练好的找矿概念模型库进行匹配,进行第一轮筛选。针对关键词匹配的结果,统计找矿概念模型库中的控矿要素,与整理好的研究区控矿要素进行匹配,计算每个控矿要素的重要性与使用率,实现找矿模型的构建与预测。

2.4 模型验证与再数据化

利用已有的找矿概念模型库,从中挑选一个找矿模型,随机对该模型进行控矿要素的删除。对模型匹配与计算之后的结果进行验证,如果该结果中存在之前被删除的控矿要素,则该结果正确可靠。通过基于深度学习找矿模型预测方法建立研究区的找矿概念模型,结合区域三维地质模型和立方体模型量化预测,对预测结果进行分析、评价,验证找矿概念模型的建立是否正确,若预测评价结果能很好地支撑区域找矿,说明找矿概念模型计算正确,可将其导入找矿概念模型数据库中,作为下次的训练样本,从而实现数据的再数据化。

3 实验与结果

3.1 资料收集

以大水金矿进行实例验证,主要流程为资料收集与找矿模型构建。资料收集是整个研究工作能

够进一步实施开展的前期基础。为建立大数据找矿模型,主要收集研究区勘查报告、研究报告、相关文献等。其中收集了勘查成果报告、区域地质资料、论文文献共计 73 篇,地质剖面图 60 条,中段平面图 22 幅,钻孔 125 个。

对收集的资料进行初步整理,步骤为:①对所有勘探线剖面图进行初步处理,结合地质知识提取出剖面图中各个地质体的分界线;②根据中段地质平面图提取出矿区的勘探线剖面工程布置图;③简单处理中段地质平面图,提取出各个地质体界线;④通过勘探线剖面图和中段地质平面图统计研究区所有地质体,包括灰岩、花岗闪长岩、闪长玢岩、方解石脉、断层破碎带,并圈定矿化蚀变带。

3.2 控矿要素分析

以成矿系列理论为指导,在已有资料及前人研究的基础上,总结区域成矿地质背景与成矿规律,归纳区域成矿系列,在区域矿化模型的基础上结合实际研究资料,构建找矿预测的地质模型。通过对研究区成矿地质背景、主成矿期、矿床成因类型、矿化类型的研究,构建同类型矿床矿化模型,并在此基础上,结合研究区实际情况,高度凝练简化,构建研究区的找矿模型。通过对研究区成矿地质背景、主成矿期次、成因类型及矿化类型的研究,基本可以确定研究区矿床的成矿地质条件、找矿标志和控矿要素。

3.2.1 深度学习模型分类

选取研究矿区关键词为“金矿”、“热液”、“低温”,带入找矿模型系统中,而后根据关键词进行匹配,在找矿概念模型库中搜寻相关矿床模型,并进行整理,之后得到矿床模型。根据关键词“金矿”,通过模型计算,从知识库中可以筛选出与“金矿”相关的找矿概念模型,例如,“中酸性、酸性浅或超浅成侵入岩有关的斑岩型金矿”、“侵入岩有关热液型金矿”、“内蒙古朱拉扎嘎金矿床”等。利用 RNN 分类算法的时间序列性,可以建立“热液”与“金矿”二者间的关系,达到更好的控矿要素匹配分类效果。例如对上述 3 个找矿概念模型进行分类,根据“金矿”“热液”,在 3 个模型中与之匹配度最高的是“侵入岩有关热液型金矿”,可归为第一大类,同时可知“侵入岩”与“热液”之间存在关联,所以在“中酸性、酸性浅或超浅成侵入岩有关的斑岩型金矿”与“内蒙古朱拉扎嘎金矿床”都只与“金矿”有关时,优先

将“中酸性、酸性浅或超浅成侵入岩有关的斑岩型金矿”归为第二大类,“内蒙古朱拉扎嘎金矿床”归为第三大类。

3.2.2 控矿要素选取

通过研究大水金矿的地质特征及控矿要素,初步确定研究区的地质找矿信息的控矿要素,并将文献资料中选取的控矿要素在系统中进行精炼、简化、提取,如图8所示。

通过与匹配的找矿概念模型中的控矿要素进行对照,实现对文献中提取的控矿要素的简化与凝练。例如针对成矿期,在文献提取中得到该地的成矿期主要为燕山期,通过多期次构造活动叠加和改造进行成矿,同时该地三叠系与成矿也密切相关,通过与匹配的找矿概念模型控矿要素,可将其简化凝练成该地成矿期为燕山期,与三叠系也密切相关。简化凝练结果如表1所示。

3.3 找矿模型构建与预测

使用提取的研究区控矿要素作为待处理数据,以找矿概念模型库为训练集,对控矿要素与找矿概念模型进行分类。对比匹配成功的模型,计算控矿要素的使用率与重要性,即对所有模型中控矿要素的出现次数与类别价值进行累加。提取相关重要性高的控矿要素,通过人机交互的方式建立大水金矿研究区的矿化模型。

表1 大水金矿控矿要素提取

Table 1 Extraction of ore-controlling factors from the Dashui gold deposit

控矿要素类型	特征参数描述
成矿地质背景	秦祁昆造山带,处在东昆仑—南秦岭褶皱带 甘肃省南部的南秦岭晚古生代—中生代多金属成矿带 金矿赋存于三叠系马热松多组的一套灰岩和白云质灰岩中
成矿期	燕山期、三叠系与成矿密切相关
含矿岩系	白云质灰岩、灰岩、闪长玢岩及花岗闪长岩
构造条件	构造运动控矿、深大断裂构造及次级断裂 深大断裂、断裂破碎带
围岩蚀变	方解石化、硅化及赤铁矿化

结合甘肃大水金矿的现有资料,在深度学习后的矿化模型基础上对大水金矿找矿模型进行研究。分析结果如表2所示。从表2可知,在构建大水找矿模型时,根据控矿要素相关重要性可知,在地球物理方面,其影响因素较多,总体相关重要性达到37,比其他方面都较高,应着重考虑地球物理方面控矿要素对模型构建的影响,同时要将地球化学、岩体条件等因素考虑在内,实现全面客观地构建大水找矿模型。

通过基于循环神经网络的找矿模型构建,自动提取大水金矿的控矿要素,构建完整的大水金矿找矿模型。通过验证可靠后,将其收录到找矿知识库



图8 系统选取研究区控矿要素

Fig. 8 Selection of ore-controlling factors in the study area

表 2 找矿模型构建

Table 2 Construction of the prospecting model

要素类型	控矿要素	特征变量	相关重要性
地层条件	大地构造位置	秦祁昆造山系	文献提取
	成矿时代	燕山期、三叠纪	
	成矿有利岩体	花岗闪长岩、闪长玢岩	
岩体条件	围岩蚀变	矿化蚀变带	13.6
	岩体缓冲区	方解石脉缓冲区、花岗闪长玢岩缓冲区、闪长玢岩缓冲区	
	断裂缓冲区	断裂缓冲区	
构造条件	构造展布特征	断裂等密度、断裂频数、断裂异常方位、断裂方位异常度	文献提取
	地球化学	Au、Ag、Sb、Fe、Cu、Au、As、Pb(Mo、Sn)晕异常及金分散流异常	13.8
地球物理	磁法	磁异常、 ΔT 高异常	12
	重力	重力正异常、重力高值异常	12
	激电	激发极化异常、高阻异常	3
	重砂异常	自然重砂异常	10

中,不断完善找矿知识库,使找矿模型构建越来越简便,预测结果更加可靠。

4 结 论

(1)通过收集已有的找矿模型,建立找矿知识库,为深度学习提供训练数据,将深度学习算法引入到找矿概念模型构建与预测中。

(2)通过对多种深度学习算法的分析研究,比较当前主流的文本分类算法,选用RNN分类方法作为模型分类算法。与其他深度学习算法相比,RNN算法可以利用更长的词序,配合Attention机制,再通过加权体矩阵体现句子中的核心词汇部位。利用RNN分类算法的Attention机制与序列记忆功能,帮助使用人员更好地对模型中的控矿要素进行评价。

(3)通过对CNN分类算法在找矿概念模型构建与预测的研究,总结出一套适用于预测研究区找矿概念模型的方法流程,提高了找矿概念模型构建与预测的工作效率。本次研究示范区依据后验概率与信息量高值区的分布情况,以后验概率值大于等于0.95、信息量值大于4.0的块体为成矿有利区域,并结合其在空间上的分布圈定了5个成矿有利区,提高了工作效率,缩短了工作时间。

(4)找矿知识库还不完善,工作程度数据库及

已知矿产地数据库仍在建设中,需要进一步收集资料,完善数据库内容。此外,本文中使用的清洗规则也不完善,有待优化。下一步工作重点是完善模型筛选后推荐算法与控矿要素评价算法,使最终找矿模型达到最优。

致谢:在找矿知识库构建与模型训练过程中得到课题组人员的大力协助,审稿专家提出了细致、中肯的建议,在此表示衷心的感谢。

参考文献

- [1]肖克炎,孙莉,李楠,等.大数据思维下的矿产资源评价[J].地质通报,2015,34(7):1266-1272.
- [2]赵鹏大,张寿庭,陈建平.危机矿山可接替资源预测评价若干问题探讨[J].成都理工大学学报(自然科学版),2004,31(2):111-117.
- [3]郑啸,李景朝,王翔,等.大数据背景下的国家地质信息服务系统建设[J].地质通报,2015,34(7):1316-1322.
- [4]赵鹏大.大数据时代数字找矿与定量评价[J].地质通报,2015,34(7):1255-1259.
- [5]周顺平,徐枫.大数据环境下地理信息产业发展的几点思考[J].地理信息世界,2014,21(1):45-50.
- [6]朱月霞.基于大数据的地质数据存储与管理研究[C]//江苏省测绘地理信息学会.江苏省测绘地理信息学会2014年学术年会论文集,2014:3.
- [7]严光生,薛群威,肖克炎,等.地质调查大数据研究的主要问题分析[J].地质通报,2015,34(7):1273-1279.
- [8]黄宜华.大数据机器学习系统研究进展[J].大数据,2015,1(1):

- 35-54.
- [9]张震,王文发.人工智能原理在人类学习中的应用[J].吉首大学学报(自然科学版),2006,(1):39-42.
- [10]张丽芳.浅谈机器学习的现状及策略[J].现代经济信息,2009,(6):155,157.
- [11]陈建平,陈勇,朱鹏飞,等.数字矿床模型及其应用——以新疆阿勒泰地区可可托海3号伟晶岩脉稀有金属隐伏矿预测为例[J].地质通报,2011,30(5):630-641.
- [12]于萍萍,陈建平,柴福山,等.基于地质大数据理念的模型驱动矿产资源定量预测[J].地质通报,2015,34(7):1333-1343.
- [13]弓小平,王世称,杨兴科,等.地质矿产预测信息化相关问题的探讨[J].地质找矿论丛,2005,(1):66-70.
- [14]Anderson T W. An Introduction to Multivariate Statistical Analysis[M]. New York: John Wiley & Sons, 2003.
- [15]Wu X, Kumar V, Ross-Quinlan J, et al. TOP 10 algorithms in data mining[J]. Knowl. Inf. Syst., 2007, 14: 1-37.
- [16]翟裕生,邓军,李晓波.区域成矿学[M].北京:地质出版社,1999.
- [17]涂光焯.超大型矿床的找寻和理论研究[J].矿产与地质,1989,(1):1-8.
- [18]朱裕生,梅燕雄.成矿模式研究的几个问题[J].地球学报,1995,(2):182-189.
- [19]程裕淇,陈毓川,毛景文,等.初论矿床的成矿系列问题[J].中国地质科学院院报,1979,(1):1-7.
- [20]李玉斌,吴静哲,陈小格.基于关键词聚类的深度学习研究[J].软件导刊,2019,18(3):199-203.
- [21]吴永亮,贾志杰,陈建平,等.基于大数据智能的找矿模型构建与预测[J].中国矿业,2017,26(9):79-84.
- [22]Li S, Chen J P, Xiang J. Prospecting Information Extraction by Text Mining Based on Convolutional Neural Networks A case study of the Lala Copper Deposit, China[J]. IEEE Access, 2018, 6: 52286-52297.
- [23]龚千健.基于循环神经网络模型的文本分类[D].华中科技大学硕士学位论文,2016.
- [24]尹汉钊.基于递归神经网络与集成算法的时间序列预测应用研究[D].西安电子科技大学硕士学位论文,2010.
- [25]林蕾.基于循环神经网络模型的遥感影像时间序列分类及变化检测方法研究[D].中国科学院大学(中国科学院遥感与数字地球研究所)博士学位论文,2018.
- [26]Alom M Z, Hasan M, Yakopcic C, et al. Inception Recurrent Convolutional Neural Network for Object Recognition[J]. arXiv: 1704.07709, 2017: 1-11.
- [27]刘腾飞,于双元,张洪涛,等.基于循环和卷积神经网络的文本分类研究[J].软件,2018,39(1):64-69.
- [28]Nallapati R, Zhai F, Zhou B. SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents[J]. AAAI, 2017, 1:1.
- [29]刘婷婷,朱文东,刘广一.基于深度学习的文本分类研究进展[J].电力信息与通信技术,2018,16(3):1-7.
- [30]Qin Y, Song D J, Chen H F, et al. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction[J]. arXiv: 1704.02971, 2017:1-7.