文章编号：2095-2163(2020)08-0118-06 中图分类号：TP391.4 文献标志码：A

基于深度学习的第三代基因测序一致性序列生成

王水介[1]，周 倩[2]，刘贤明[1]
（1 哈尔滨工业大学 计算机科学与技术学院，哈尔滨 150001；2 鹏城实验室，深圳 518040）

摘 要：继人类基因组计划开展以来，基因测序已经广泛影响了生命科学的研究方式。基因组组装是从大量随机测序获得的短片段中重建出基因组长序列的过程，其最终目标是生成完整、准确的一致性序列，为后续多种研究提供可靠的参考基因组。第三代基因测序技术可以产生读长达几十 kb 的片段，其应用极大提高了基因组组装的完整性，但测序的高错误率却限制了最终一致性序列的准确性。本研究提出基于深度学习的一致性序列生成模型，利用人工神经网络提取基因多序列比对结果的结构特征，生成准确率更高的一致性序列。实验表明，该模型针对第三代测序数据可以生成质量较高的一致性序列，并且无需读取测序时的质量值，也不用一次读取超长序列，可以更加灵活地处理小数据块。

关键词：第三代基因测序；一致性序列；深度学习；人工神经网络

DLCC：Deep-Learning-Based Consensus Construction from long error-prone reads

WANG Shuijie[1]，ZHOU Qian[2]，LIU Xianming[1]
（1 College of Computer Science and Technology，Harbin Institute of Technology，Harbin 150001，China；
2 Pengcheng Laboratory，Shenzhen 518040，China）

【Abstract】 Since the launch of the Human Genome Project，genome sequencing has widely influenced the way life sciences are studied. Genome assembly is the process of reconstructing long genome sequences from a large number of short fragments obtained by random sequencing. Its ultimate goal is to generate complete，accurate consensus，providing reliable reference genome for subsequent analyses. The application of the third-generation sequencing technology，who can generate the read as long as dozens of Kb，has greatly improved the integrity of genome assembly，but the high error rate of the long read limits the accuracy of the final consensus sequences. This study proposes a consistent sequence generation model based on deep learning，which uses artificial neural network to extract the structural characteristics of gene multiple sequence alignment results to generate consensus with higher accuracy. Experiments show that the model can generate high-quality consensus for the third-generation sequencing data，while it does not need to read the quality value during sequencing，nor to read ultra-long sequences at a time，so it can process small data blocks more flexibly.

【Key words】 the third-generation sequencing；consensus；deep learning；artificial neural network

0 引 言

基因测序是一种新型基因检测技术，能够从血液或唾液中分析测定基因全序列，来预测罹患多种疾病的可能性、个体的行为特征及行为。基因测序技术能锁定个人病变基因，提前预防和治疗。基因测序相关产品和技术已由实验室研究演变到临床应用，可以说基因测序技术，是下一个改变世界的技术[1]。2020 年初，随着新型冠状病毒肺炎疫情的爆发，针对病毒基因组分析得到了更多的重视，快速获得新冠病毒基因组的参考序列是病毒核酸检测和生产疫苗的基础[2]。对多个病毒基因组和中间宿主动物基因组的测序和比对，可以了解病毒的来源、感染物种和在自然界中的变异情况，以此估计病毒的传播难易程度，从而为控制疫情的蔓延提供理论指导[3]。

利用第三代测序片段进行基因组组装、变异检测等已经成为基因组学领域的基本分析手段。其测序基因组覆盖均匀、长读长的优势极大提高了基因组的组装的完整度和连续性。然而其读长（1 kb-50 kb）和高错误率（~15%）对组装过程中的序列比对产生了极大挑战，也影响了最终组装出的参考基因组序列的准确性[4]。目前基于第三代测序序列得到的组装、分析结果仍然需要利用准确率较高的第二代测序数据进行校正[5]。因此，对第三代测序数据进行抛光的重要性不言而喻，这样能够快速高效地对其进行部分错误纠正，并生成准确率较高的一致性序列，对长序列的组装以及后续的基因组学研究都有巨大的意义[6]。

1 第三代测序数据预处理

引入深度学习方法进行一致性序列生成任务，

作者简介：王水介（1996-），男，硕士，主要研究方向：人工智能、第三代测序；周 倩（1991-），女，博士，主要研究方向：基因组数据处理；刘贤明（1983-），男，博士，教授，主要研究方向：人工智能、图像处理、基因组数据处理。

收稿日期：2020-06-17

首先要对原始数据预处理，使其适合作为人工神经网络的输入。本研究中采用的数据集为公共数据库中的 Oxford Nanopore Technology 公司的纳米孔测序数据（即第三代基因测序 ONT 数据），选取的模式物种包括大肠杆菌、酵母菌以及果蝇，数据集的具体信息见表 1。

表 1　实验所用 ONT 数据统计

Tab. 1　The statistics of the ONT data used in the experiments

ONT 序列统计[a]	E.coli.	Yeast	Fly
序列数量	34 438	146 967	663 784
平均长度	6 710	11 038	66 956
最大长度	28 385	414 134	446 050
最小长度	276	6 000	5
碱基总数	231 078 830	1 622 148 184	4 617 586 601
测序深度	54×	134×	34×

[a]数据来源：E.coli 和 Fly 数据下载自 NCBI（https://www.ncbi.nlm.nih.gov/sra），索引号分别是 ERR1147230 和 SRR6702603，Yeast 数据下载自　http://www.genoscope.cns.fr/externe/Download/Projets/yeast/datasets/raw_data/S288C/

3 个模式物种的参考基因组长度分别为 4,641,652 个碱基、12,071,326 个碱基、137,547,960 个碱基。由于这 3 个模式物种的全基因组长度、复杂程度及杂合情况均有较大差异，可以较好地评估本文提出方法的完整性和全面性。酵母菌参考基因组序列的片段示例如图 1 所示。

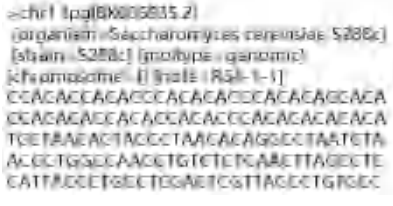

图 1　酵母参考基因组染色体 1 序列片段

Fig. 1　The chromosome 1 fragment of Yeast reference genome

经过不断试验，对第三代测序原始数据采用如下处理方式：首先，测序过程不同通量数据并非是完全对齐，而是呈阶梯状排布，相邻两条之间都有一定碱基数的错位，因此需要将测序序列比对到初始参考基因组骨架上，确定测序序列之间的排列顺序，此步骤通过快速比对软件 miniasm2 完成；其次，裁剪比对后的结果，使大部分序列可以头尾对齐。以深度为 40×的数据块为例，具体做法是取正向第十五条序列的尾部向前 50 个位点作为块截止位点，反向第十五条序列头部向后 50 个位点作为块起始位点；最后，将裁剪后的比对结果分割为深度为序列乘数、宽度为 12 个位点的小块，每一个小块用于预测块中心 4 个位点的碱基种类，将每个小块的预测结果最终拼接成完整的一致性序列。

2　一致性序列生成模型

一致性序列生成任务是通过高通量的序列计算得到的，并且基因序列中存在一定的结构相关性，即序列某一位点前后的碱基对此位点碱基的预测会起到影响作用，这些特征从直觉上符合人工神经网络适用的场景。最终的实验结果论证了在一条序列中以及多条序列间均存在相关性，也证明了采用深度学习方法的可靠性。

具体采用的网络结构如图 2 所示。主要分为卷积神经网络模块、通道注意力机制模块、循环神经网络模块、多任务学习模块四部分。四个结构相结合，可以有效地提取出一条序列内部的结构相关性以及不同通量序列之间的测序时序上的相关性，还可以赋予对一致性序列影响较大的部分序列更大的权重，将提取到的不同层次相关性充分利用，以得到更优秀的结果。

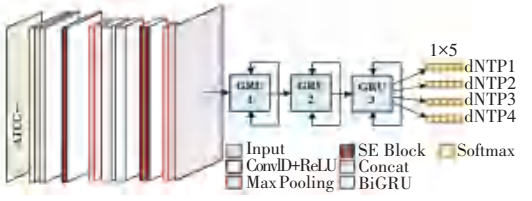

图 2　一致性序列生成网络结构图

Fig. 2　Consensus construction network

网络中第一部分是卷积模块，卷积神经网络（Convolutional Neural Network，CNN）是深度学习中一种被广泛应用的网络结构，在深度学习技术快速发展的背景之下，首先在计算机视觉领域收获了卓越的成绩[7]。在一致性序列生成任务中，由于输入到卷积模块中的数据是切割成固定尺寸的第三代测序数据比对块，其通量和每条序列包含的位点个数都是确定的。分析比对块的特点可知，这是行数代表序列通量、列数代表每条序列位点数的碱基矩阵。通过合理推测以及实验证明，不同序列可以看作不同时序的信号，用卷积结构对其进行特征提取的效果并不理想。结合上述的分析，由于高通量的基因序列数据不同通量之间有时序关系，经过对比发现更适合使用循环网络来利用其相关性。因此，卷积模块中采用 1×3 的一维卷积核，仅仅用作横向的特征提取，即提取一条序列中碱基之间的相关性特征，此做法不仅更合理地利用了数据的特性进行模型设计，还可以减少参数量，使得运算效率大大增加，在

更短时间获得质量更高的结果,具体卷积的形式如图 3 所示。同样的,也采用 1×2 的平均池化,用于减少参数量,有效地压缩数据以及参数的规模,减小计算的复杂性,降低时间代价。此外,考虑到如果采用平均池化的方式,可能会使得对某个位点碱基种类影响最大的碱基信息变得模糊,受到其他重要程度更低的碱基信号影响,使得最终结果不能达到令人满意的准确率,因此决定在模型中采用最大池化的形式。

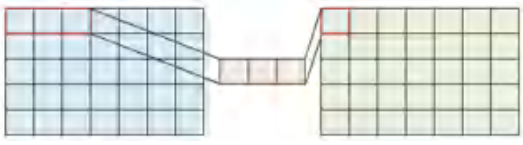

图 3 一维卷积提取水平特征

Fig. 1 One dimensional convolution to extract lateral features

在一致性序列预测任务中,预测的碱基在块中的位置基本确定,但是需要更多地利用周围碱基的信息,以纠正测序过程中产生的错误,而非仅仅考虑对应位置的碱基种类。通道注意力机制是通过引入通道之间的相关性,而并不会只着重关注部分碱基情况。模型中采用 Squeeze-and-Excitation Block 的结构,首先是 Squeeze 操作,顺着空间维度特征压缩,将每个二维的特征通道变成一个实数,这个实数某种程度上具有全局的感受野,并且输出的维度和输入的特征通道数相匹配,表征着在特征通道上响应的全局分布,而且使得靠近输入的层也可以获得全局的感受野[8];其次是 Excitation 操作,是一个类似于循环神经网络中门的机制,通过参数来为每个特征通道生成权重,其中参数被学习用来显式地建模特征通道间的相关性。在全局平均池化(Global-Average-Pooling)之后有两个全连接层(Fully-Connected),分别应用 ReLU 以及 Sigmoid 激活函数,具有更多的非线性,可以更好地拟合通道间复杂的相关性,极大地减少了参数量和计算量。通过一个 Sigmoid 的门,获得 0~1 之间归一化的权重,最后将归一化后的权重加权到每个通道的特征上。

提取出横向特征并赋予通道注意力后,将会输入循环神经网络模块进行不同通量间的特征提取。循环神经网络(Recurrent Neural Network,RNN)同卷积神经网络一样,是深度学习中一种重要的网络结构,其目的是对时序信号有效的处理。循环神经网络的特点在于考虑了原始数据中的时序相关性,在这个网络结构中每一时刻的输出不仅仅与当前的输入有关,还与之前所有时刻的输入有着很大的联系,这样侧重时间相关性特点的网络结构,对于处理有着明显时序性特征的信号处理任务有着巨大的帮助[9]。一个经典的循环神经网络基本结构如图 4 所示。

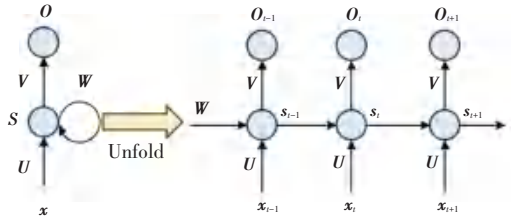

图 4 循环神经网络

Fig. 4 Recurrent neural network

在一致性序列生成模型中,循环神经网络部分采用三层的双向 GRU 结构,GRU 是 LSTM 的一种变体,相比于 LSTM 的 3 个门(forget,input,output),GRU 只有两个门(update 和 reset)。综合来看两者的性能在很多任务上不分伯仲,但是 GRU 参数相对少,更容易收敛,同时可以一定程度地避免梯度消失的问题,更适用于本课题的任务。而双向 GRU 是将前向 GRU 与后向 GRU 结合,用于解决单向的结构无法编码从后向前序列信息的问题,这样可以捕捉更全面的不同通量序列之间的语义依赖[10]。

此模块主要分为三部分:通量信息提取层、通量信息表示层、特定位点预测层。

(1)通量信息提取层:由卷积提取的各个通道特征进行列维度的拼接后,得到通量信息矩阵,作为下一步的输入;

(2)通量信息表示层:由于不同通量的序列可以视作是不同时序的结果,考虑到测序过程前后两个方向可能均包含时序上的信息,因此使用三层双向的 GRU 结构对通量信息建模,从前向后的编码信息与从后到前的编码信息相结合,此双向的编码结构作为通量信息的表示;

(3)特定位点预测层:由一层感知器结构的全连接层完成,作为某个位点碱基种类的预测方式。

在循环网络之后,加入多任务的结构,即对一个输入块最终将输出 4 个位点的预测结果,多个任务同时预测时,减少了数据来源的数量以及整体模型参数的规模,使预测更加高效。4 个任务相关性较强,经过实验验证,可以有效提升一致性序列预测的准确率。

3 实验设置及结果分析

本研究在酵母菌、大肠杆菌、果蝇 3 个模式物种

的纳米孔单分子测序数据上从多个指标来评价所提出一致性序列生成模型的性能，这 3 个物种的 ONT 序列长度、染色体条数、复杂程度以及杂合程度均有不同，因此可以不同角度考量模型的可靠程度，使得评价结果不具有特殊性或偏好性。

对最终一致性序列质量的评价指标，最直观也是最重要的一项就是通过与已发布的“金标准”参考基因组进行序列比较，计算与整个参考基因组匹配的碱基数目和预测准确率。但是由于组装过程中可能存在组装不完整、大片段缺失、倒位等问题，引起一致性序列与参考基因组比对时的困难，因此仅依靠匹配数来衡量准确率不能完全展示抛光后的一致性序列的质量。参考本领域内其他软件的评价方法，本文考虑以如下四点标准衡量该模型：

（1）覆盖参考基因组的位点数（cover reference），用于评价生成一致性序列的完整度，即从第一个匹配位点到最后一个匹配位点之间的数目，不包含长段的缺失，但是包含插入、小段缺失以及替换等错误。

（2）产生的替换数目，即发生在某个位点上一致性序列中的碱基与参考基因组同一位置的碱基不同，但是产生不同的原因也可能是此位点存在杂合情况，由于出现的概率不大，因此在统计中均计入替换数目。

（3）产生的缺失数目，这也是评判一致性序列质量时主要关注的问题。小数目的缺失主要是由第三代测序原理的特性造成的。当序列中出现连续的重复碱基时，测序仪对荧光信号或者电信号的解读精确性不够，导致转换为碱基信号时容易缺失碱基，这也是第三代测序技术准确率提升的最大瓶颈。这个缺陷在抛光过程中可以被部分地解决，因此在评价方法有效性时，对不通长度区间的缺失进行统计，长度单位为碱基对（bp），主要分为 1 bp、2 bp、3-50 bp、50-1 000 bp 这 4 个长度范围。特别地，由于果蝇的全基因组包含碱基数目更为庞大，产生的错误也更严重，因此还需统计长度大于 1 000 bp 的缺失。

（4）产生的插入数目，与缺失数目的统计方法类似，分为 1 bp、2 bp、3-50 bp、50-1,000 bp 这 4 个长度范围进行计算。同样地，对于果蝇数据，还将统计其长度大于 1 000 bp 的插入。

本文的实验覆盖了酵母、大肠杆菌、果蝇这 3 个模式物种的全基因组纳米孔测序数据，可以对模型在不同结构特征染色体上的性能全面地评估。针对酵母基因组，增加了一组在不同数据量上的实验，以研究该模型在不同测序深度上的性能。通过以上实验可以清晰地得知本文提出方法在不同物种、不同规模数据、不同裁剪尺寸上的实际效果。作为比较，对比实验将采用目前主流的具有一致性序列生成功能且结果优秀的软件进行，包括 Wtdbg、Canu、Flye 以及 Racon。

相比于目前主流的基于第三代测序数据的一致性序列生成或者基因组装抛光软件，本文提出的基于深度学习方法的模型在准确率上有明显提高，在参考基因组覆盖长度、替换数目、缺失数目、插入数目四项指标上均一定程度地表现出优势，也证明了算法的有效性及可靠性。

表 2 记录了采用不同方法对酵母纳米孔测序数据抛光的效果对比。与其他软件生成的一致性序列相比较，本文提出的方法在覆盖参考基因组的位点数上有明显的优势，同时替换以及短缺失问题均有明显减少。然而，由于在处理原始数据时为了尽量减少大量的空位（gap）以及减少含有重复片段的序列，因此导致结果中大片段的插入数量稍多。但从插入的位点数而言，却依旧保持了较好的效果，所生成的一致性序列质量有明显提升。

表 3 记录了采用不同方法对大肠杆菌纳米孔测序数据抛光的效果对比。大肠杆菌基因组结构简单，基因组数据长度较短，因此测序和组装过程中得到的序列质量较高，最终生成的一致性序列也都有很高的准确率。在具体实验中，本文提出方法在参考基因组覆盖率上已达到百分之百，并且在替换、短缺失两项依然保持着优秀的效果。此外，数据预处理的方式导致长段插入略微增加，但各项指标依旧有着不错的结果。

表 4 记录了采用不同方法对果蝇纳米孔测序数据抛光的效果对比。果蝇染色体蕴含的信息相比于前两个模式物种复杂程度有极大的提升，不仅长度的量级有所增加，也有更多杂合位点以及更难处理的结构信息。不同组装软件构建的果蝇的基因组结果质量参差不齐，大片段错误较多，为后续一致性序列的生成、抛光带来了非常大的挑战。本文提出的模型相比于其他方法依然保持了不错的效果，在参考基因组覆盖率、缺失、替换、插入等各项指标均名列前茅，但不可避免地由于数据本身以及数据预处理的特点，长段缺失较多，且第四条染色体的碱基发生了大量的缺失。总的来说，虽然最终结果有一些不足，但是基于深度学习的模型仍然提供了令人满意的一致性序列生成结果。

表 2 评估不同方法对酵母纳米孔测序数据抛光的效果对比

Tab. 2 Assessment of consensus sequences of yeast genome

Yeast 130×	Wtdbg	Canu	Flye	Racon	DLCC
参考基因组长度	12 071 326	12 071 326	12 071 326	12 071 326	12 071 326
Cover Ref	11 652 075	11 871 961	11 803 670	11 647 840	12 041 917
替换	11 521	11 150	21 724	31 077	11 164
1 bp 缺失	66 975	90 136	66 301	48 586	43 872
2 bp 缺失	5293	12743	7 114	4 732	4 960
3-50 bp 缺失	1 435	7 640	4 870	3 293	2 489
50-1 000 bp 缺失	3	3	1	3	5
1 bp 插入	10 075	1 425	8 919	10 178	5 869
2 bp 插入	4 731	699	1 452	9 705	1 036
3-50 bp 插入	5 984	626	636	26 568	2 173
50-1 000 bp 插入	4	7	8	7	164

表 3 评估不同方法对大肠杆菌纳米孔测序数据抛光的效果对比

Tab. 3 Assessment of consensus sequences of E.coli genome

E.coli 54×	Wtdbg	Canu	Flye	Racon	DLCC
参考基因组长度	4 641 652	4 641 652	4 641 652	4 641 652	4 641 652
Cover Ref	4 619 875	4 621 410	4 633 681	4 640 236	4 641 652
替换	1 210	248	376	847	433
1 bp 缺失	19 303	15 028	11 793	12 725	10 043
2 bp 缺失	4 873	4 642	2 882	3 046	2 826
3-50 bp 缺失	922	907	438	452	538
50-1000 bp 缺失	1	1	1	1	1
1 bp 插入	928	53	1 352	1 483	1 189
2 bp 插入	48	3	72	575	55
3-50 bp 插入	9	1	19	489	11
50-1 000 bp 插入	2	2	2	2	3

表 4 评估不同方法对果蝇纳米孔测序数据抛光的效果对比

Tab. 4 Assessment of consensus sequences of fly genome

Fly 32×	Wtdbg[17]	Canu[18]	Flye[19]	Racon[20]	DLCC
Ref 长度	137 547 960	137 547 960	137 547 960	137 547 960	137 547 960
Cover Ref	122 684 667	128 546 930	130 271 254	128 374 616	128 445 616
替换	179 646	96 143	110 768	199 503	98 422
1 bp 缺失	1 147 336	755 774	314 651	384 080	307 349
2 bp 缺失	164 971	122 632	24 667	43 917	32 768
3-50 bp 缺失	40 069	48 561	7 097	9 888	24 561
50-1 000 bp 缺失	26	74	48	1 102	56
1 bp 插入	59 743	23 190	100 961	249 834	43 937
2 bp 插入	3 577	1 495	2 656	75 667	2 238
3-50 bp 插入	1 502	910	1 296	31 925	1 433
50-1 000 bp 插入	41	86	141	167	66

不同物种上的实验结果可以证明基于多任务神经网络模型可以适用于不同结构特征的基因数据，并取得不错的效果。然而数据深度较高时，计算规模也会极大，因此也希望考察模型在低深度数据上是否能够取得令人满意的结果，即在同一物种不同通量的数据上进行实验。

目前也在酵母 ONT 数据上测试了使用不同深度数据的准确率变化情况，具体结果见表 5。发现在从较低乘数提升至 40×时，准确率上升快，但是超过 50×后，准确率提升并不明显，且计算代价有较大增加，使得训练难度变大。虽然在基于深度学习的模型上使用较低乘数数据生成的一致性序列准确率不如使用高乘数数据完成相同任务的结果，但是依然能够保持一个较高的数值，可见神经网络对于特征的提取相比传统方法更加优异，这也为该方向的后续研究提供了更多思路。

表 5　不同乘数酵母菌纳米孔单分子测序数据训练准确率对比

Tab. 5　Comparison of training accuracy of yeastONT data at different depths

物种	乘数(N×)	准确率/%
Yeast	20×	97.26
	30×	97.53
	40×	97.92
	50×	97.98
	59×	98.26

4　结束语

本文引入深度学习的方法提取多序列比对结果的结构特征，以生成准确率更高的一致性序列，针对低乘数的数据可以保持良好效果，并且无需读取测序时的质量值。该模型合理利用序列的局部相关性以及高通量数据中蕴含的时序信息，并将这些特征尽可能放大，显著提升了最终的一致性序列生成结果。实验结果表明，通过卷积神经网络结构提取的单序列的局部结构信息，以及循环神经网络提取的不同通量序列之间的时序信息，对最终的预测结果都有明显的提升。单序列信息提取、通量信息提取、通量信息表示、特定位点预测这样的网络分层结构，也取得了显著的效果。虽然模型中不可避免地存在一些不足之处，但是对其有效性和可靠性并不能产生影响。此外，利用人工神经网络模型完成一致性序列生成任务也开创了此方向的先河，为后续的研究工作提供了更多思路。

参考文献

[1] METZKER M L . Sequencing technologies - the next generation [J]. Nature Reviews Genetics, 2010, 11(1):31-46.

[2] WU F, ZHAO S, YU B, et al. A new coronavirus associated with human respiratory disease in China [J]. Nature, 2020, 579 (7798):1-8.

[3] XU X, CHEN P, WANG J, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission [J]. Science China Life Sciences, 2020, 63(3): 457-460.

[4] YE C , MA Z S. Sparc: A sparsity-based consensus algorithm for long erroneous sequencing reads[J]. PeerJ, 2016, 4(24):e2016.

[5] LOOSE M, MALLA S, STOUT M. Real - time selective sequencing using nanopore technology[J]. Nature methods, 2016, 13(9): 751-754.

[6] ISTACE B, FRIEDRICH A, D´AGATA L, et al. de novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer [J]. Gigascience, 2017, 6(2): giw018.

[7] LI H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences [J]. Bioinformatics, 2016, 32 (14): 2103-2110.

[8] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient - Based Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[9] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[10] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder - decoder for statistical machine translation [J]. arXiv preprint arXiv: 1406. 1078, 2014.

(上接第 117 页)

[10] LI Cuiping, QIN Jiexuan, LI Jiajie, et al. The accident early warning system for iron and steel enterprises based on combination weighting and Grey Prediction Model GM (1, 1) [J] . Safety Science,2016,89.

[11]韩泉叶,王晓明,党建武. 城市轨道交通线网突发应急事件分类分级模型研究[J]. 城市轨道交通研究,2011,14(10):37-40.

[12] ZHANG Dongming, DU Fei, HUANG Hongwei. Resiliency assessment of urban rail transit networks: Shanghai metro as an example[J]. Safety Science,2018,106.

[13]花玲玲,郑伟. 基于复杂网络理论的铁路事故致因分析[J]. 中国安全科学学报,2019,29(S1):114-119.

[14]MA Jiaqi, DAI Hong. A methodology to construct warning index system for coal mine safety based on collaborative management [J]. Safety Science,2017,93.

[15]张振宇. 基于安全域的轨道交通路网安全状态评估与预测[D]. 北京交通大学,2016.

[16]温念慈,倪少权,陈钉均,等. 城市轨道交通突发大客流协同应急决策研究[J]. 中国安全生产科学技术,2017,13(7):48-54.

[17]蔡正杰,梁昌勇,赵树平. 突发环境事件等级评估方法研究[J]. 计算机应用研究,2014,31(11):3217-3220.

[18]候晓丽. 城市轨道交通运营安全风险预控的研究[D]. 北京交通大学,2017.