

## 基于BiLSTM-CRF的航行通告命名实体识别研究

项恒, 杨明友, 李猛

引用本文

项恒, 杨明友, 李猛. 基于BiLSTM-CRF的航行通告命名实体识别研究[J]. 计算机科学, 2024, 51(11A): 240300148-6.

XIANG Heng, YANG Mingyou, LI Meng. Study on Named Entity Recognition of NOTAM Based on BiLSTM-CRF [J]. Computer Science, 2024, 51(11A): 240300148-6.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多模态融合的动态恶意软件检测方法](#)

Multimodal Fusion Based Dynamic Malware Detection

计算机科学, 2024, 51(11A): 240200098-7. <https://doi.org/10.11896/jsjcx.240200098>

[基于开放集的入侵检测方法研究](#)

Study on Open Set Based Intrusion Detection Method

计算机科学, 2024, 51(11A): 231000033-6. <https://doi.org/10.11896/jsjcx.231000033>

[基于CNN结合BiGRU的恶意流量分类算法研究](#)

Study on Malicious Traffic Classification Algorithm Based on CNN Combined with BiGRU

计算机科学, 2024, 51(11A): 231100106-9. <https://doi.org/10.11896/jsjcx.231100106>

[基于深度学习智能反射面辅助通信系统的联合波束成形](#)

Deep Learning Based Joint Beamforming in Intelligent Reflecting Surface Enhanced Wireless Communication Systems

计算机科学, 2024, 51(11A): 231200125-5. <https://doi.org/10.11896/jsjcx.231200125>

[基于因果关系的领域泛化长尾学习](#)

Domain Generalization and Long-tailed Learning Based on Causal Relationships

计算机科学, 2024, 51(11A): 240300041-8. <https://doi.org/10.11896/jsjcx.240300041>

# 基于 BiLSTM-CRF 的航行通告命名实体识别研究

项恒 杨明友 李猛

中国民航大学空中交通管理学院 天津 300300

**摘要** 针对当前国际民航组织对数字航行通告研究仅考虑对文本航行通告环境兼容,而未考虑对数字航行通告环境兼容的问题,提出一种基于 BiLSTM-CRF 的航行通告命名实体识别模型,以实现文本航行通告中相关实体的自动识别,并为转换数字航行通告提供所需的基本数据。通过构建航行通告语料标记数据集对 LSTM, BiLSTM, BiLSTM-CRF 3 种模型进行对比实验。实验结果显示,所提模型的精确率、召回率、 $F_1$  值分别为 95%, 95%, 95%, 验证了其在航行通告领域的有效性,证明本研究可以有效识别航行通告中的重要实体信息。

**关键词** 航行通告;命名实体识别;深度学习;双向长短期记忆网络;条件随机场

**中图分类号** V355; TP391

## Study on Named Entity Recognition of NOTAM Based on BiLSTM-CRF

XIANG Heng, YANG Mingyou and LI Meng

College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China

**Abstract** Aiming at the problem that the current research of International Civil Aviation Organization in digital NOTAMs, which only considers the compatibility with the environment of textual NOTAMs, but not digital NOTAMs, a named entity recognition model for NOTAMs based on BiLSTM-CRF is proposed to realise the automatic recognition of relevant entities in textual NOTAMs and to provide the necessary basic data for the conversion of digital NOTAMs. Comparative experiments are carried out by constructing a NOTAM corpus tagged dataset in three models, LSTM, BiLSTM and BiLSTM-CRF, and the experimental results show that the precision, recall and  $F_1$  value of the proposed method is 95%, 95% and 95%, respectively, which verifies the effectiveness of the proposed method in the field of NOTAMs and proves that this study can effectively obtain the important entity information in NOTAMs.

**Keywords** NOTAM, Name entity recognition, Deep learning, BiLSTM, CRF

### 1 引言

航行通告<sup>[1]</sup>是一种以航空固定业务通信网发布的,告知飞行员和与飞行运行有关的人员关于航空设施、服务、程序的设立、情况或者变化,以及对航空有危险的情况出现和变化的通知。因受传输网路限制,航行通告仅能以文本形式进行传输。

一份传统文本航行通告主要由电报报头、系列编号和航行标志、限定行 Q 项、发生地 A 项、生效时间 B 项、失效时间 C 项、分段时间 D 项、航行通告正文 E 项、下限 F 项和上限 G 项组成。其示例片段如图 1 下半部分所示,除 E 项外,其他数据项都以结构化呈现,而 E 项则以明语和简缩字填写,结构松散,难以解析,需掌握相应航空专业词汇和书面用语方可读懂。

随着民航业的迅速发展,文本形式的航行通告已无法满足当前的运行环境,现面临以下挑战:拍发数量的激增,致使当前航行通告处理系统无法高效地对通告进行过滤,飞行前资料公告信息冗余,增加飞行人员的航前准备负担;文本形式结构且大量缩略语的使用,使航行通告可读性差;航行通告

质量需人工进行审核,耗时耗力,数据质量无法得到保证。针对该问题,国际民航组织(International Civil Aviation Organization, ICAO)提出了“数字航行通告”概念。数字航行通告<sup>[2]</sup>是一种以计算机可处理的形式来定义文本航行通告所规定需要发布信息的小型数据集,以数字形式发布航行通告,相关处理系统能够更加高效地对其进行处理,必要时,可对含有地理信息的数据进行可视化实时动态展示,使飞行员或运行相关人员能够更加直观地了解机场或空域的运行状态,增强情景意识,保障航班安全高效运行。

ICAO 倡议,未来数字航行通告应以渐进式方式实施。因此,在未来实际运行环境中,数字航行通告与传统文本航行通告将并行发布。而当前 ICAO 对数字航行通告的研究仅考虑到兼容文本航行通告运行环境,而并未考虑到文本航行通告兼容数字运行环境的问题,即能否将文本航行通告转换为数字航行通告。对于纯数字环境运行的应用程序来说,全数字数据处理能够高效利用应用程序,反之则增加应用程序的负担。因此,就有必要对文本航行通告如何转换为数字航行通告进行研究。

航行通告场景复杂多样,不同场景对应不同的基本数据。

基金项目:国家自然科学基金(U1833103)。

This work was supported by the National Natural Science Foundation of China(U1833103)。

通信作者:项恒(5136849@qq.com)

为解决上述问题,本文以航行通告跑道关闭场景为例,将模型简化为单一航行通告场景下的命名实体识别的问题。在对跑道关闭场景文本航行通告数据结构与数字航行通告所需基本数据表单研究分析的基础上发现,其跑道关闭场景文本航行通告与数字航行通告所需基本数据表单有对应关系,如图 1 所示。

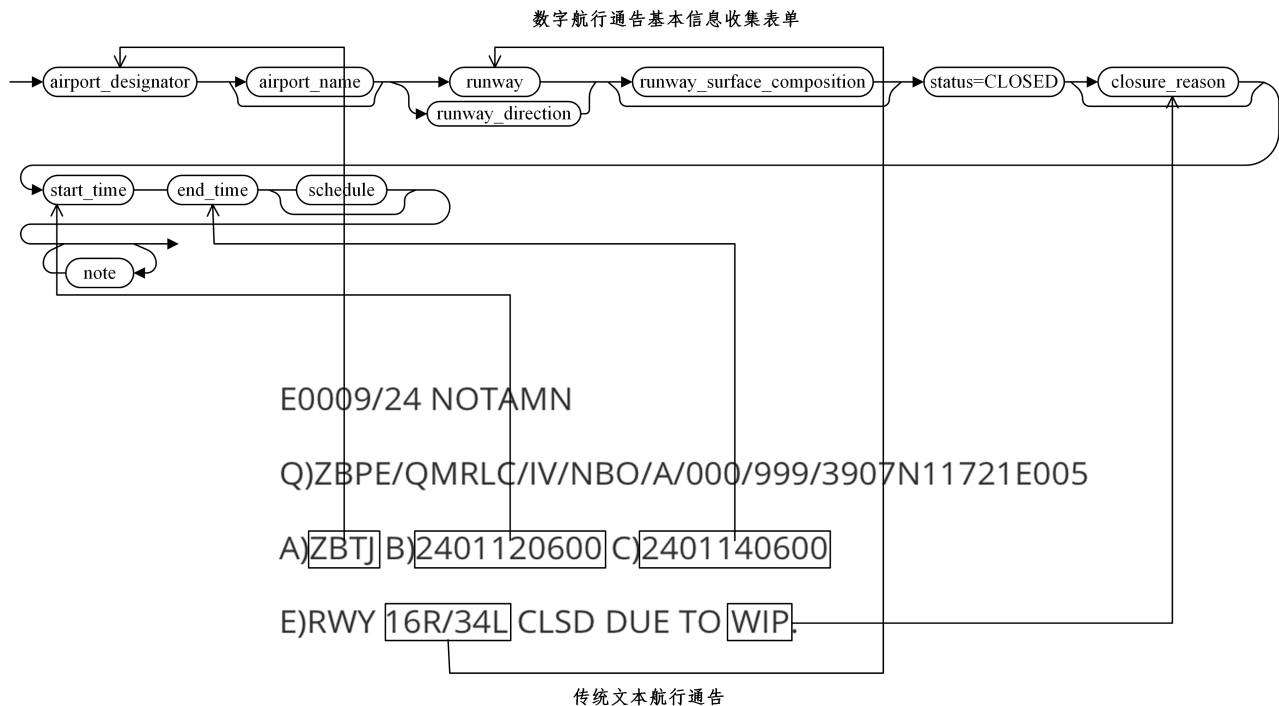


图 1 文本航行通告与数字航行通告基本数据表单对应关系

Fig. 1 Correspondence between basic data sheets for text and digital NOTAMS

## 2 相关工作

在自然语言处理领域,命名实体识别(Name Entity Recognition,NER)被视为一项基本的工作,是信息抽取的第一个关键环节,为其他许多下游任务的高效完成奠定了基础。NER的演进可概括为3个主要阶段:依赖规则及词典的技术、采用统计机器学习技术和运用深度学习技术。

NER早期阶段,基于规则的方法在特定领域的文本上展现了较好的识别性能,但它同时面临着规则制定成本高昂以及规则难以适用于不同文本集的限制,一旦更换领域,就需要花费大量时间来对其规则和模式进行重构<sup>[3-4]</sup>。

NER中期阶段,随着机器学习在自然语言处理领域的兴起,为弥补基于规则和词典方法的缺点,研究领域逐步倾向于探索机器学习技术在NER中的应用,旨在一定程度上克服基于规则及词典方法的缺陷。基于统计机器学习的方法是通过建立概率模型将实体识别转化为参数概率的分类问题,其主要方法包括:隐马尔可夫模型(Hidden Markov Model, HMM)<sup>[5]</sup>、最大熵(Maximum Entropy Model, MEM)<sup>[6]</sup>、条件随机场(Conditional Random Fields, CRF)<sup>[7]</sup>、支持向量机(Support Vector Machine, SVM)<sup>[8]</sup>等。Bikel等<sup>[9]</sup>采用HMM来分析文本,通过考虑大小写、数字和句首词等特征,来计算文中人名、日期和数值等实体出现的概率。Brothwick等<sup>[10]</sup>在综合分析首字母大写、句末信息及文本是否为标题等多重特征后,将MEM应用于NER任务。Mccallum等<sup>[11]</sup>提

出基于CRF对英文及德文进行实体抽取的方法。Takeuchi等<sup>[12]</sup>采用SVM对MUC-6评测数据集和分子生物学相关文本进行实体识别分析,结果显示该模型在生物科学领域的命名实体识别中展现出了优秀的性能。

NER近期阶段,实体识别主要利用深度学习(Deep Learning)。与统计机器学习相比较,深度学习在实体抽取方面能够充分利用神经网络发现并提取文本中的隐藏特征,显著减少复杂特征工程的构建和降低对高质量数据语料的依赖性。处理NER任务的常见组合为深度学习模型-CRF,其中深度学习包括CNN(Convolutional Neural Network, CNN), RNN(Recurrent Neural Network, RNN)和GNN(Graph Neural Network, GNN)等。Collobert等<sup>[13]</sup>提出一种基于CNN的实体识别方法,该技术免除了手动特征输入的需求,利用大规模未标注的训练数据学习内在表示,减少了输入阶段的特征预处理需求,并采用了端到端训练的深层神经网络结构。然而,CNN模型存在不能有效记忆上下文信息的限制,为克服这一缺陷,Huang等<sup>[14]</sup>提出了基于LSTM的序列标注模型,包括LSTM, Bi-LSTM和Bi-LSTM-CRF的结合。Bi-LSTM-CRF模型特别强调了上下文信息的利用,并且能够接受整个句子作为输入。Cetoli等<sup>[15]</sup>在NER任务中首次使用GCN来解决实体识别问题,在BiLSTM-CRF层中间加入一层GCN层,该层利用句子句法依存关系构图,通过GCN将节点信息传递给附近节点,将N层图堆叠在一起,该网络结构可传播最多相距N跳的节点特征。随着计算机技术的

快速发展,当前 NER 领域涌现出越来越多的深度学习模型,如谷歌 Devlin 等<sup>[16]</sup>提出的一种新的语言表示模型,被称为 BERT。该模型提升了许多下游自然语言处理任务的效果,研究者只需在原模型基础上进行微调,就能训练出相关任务的最先进模型。如 Guo 等<sup>[17]</sup>为对法律文本中的案件实体进行智能识别,提高案件的处理效率,提出一种基于 BERT-BiLSTM-CRF 模型,并通过实验验证了该模型的性能。Liu 等<sup>[18]</sup>总结了当前 BERT 模型优化改进的主要趋势和方向。建议未来研究可以重点关注目标优化和 Transformer 结构改进方面,并考虑不同应用场景下的特定需求。

对于文本航行通告 NER, Li 等<sup>[19]</sup>从文本航行通告角度,基于机器学习实现了对航行通告进行文本分类。Xiang 等<sup>[20]</sup>基于自然语言处理,对不规范文本航行通告进行识别。Pan

等<sup>[21]</sup>通过令牌化实现分词,并通过词嵌入方法,实现了航行通告命名实体抽取。上述研究主要集中于传统机器学习的方法对文本航行通告进行处理,并未涉及深度学习领域在文本航行通告中的应用。对此,本文提出的模型标志着深度学习在此领域的首次应用。

### 3 BiLSTM-CRF 模型构建

#### 3.1 模型整体架构

本文构建了一种用于识别文本航行通告中特定实体的模型结构,如图 2 所示。该模型涵盖了嵌入层、BiLSTM 层以及 CRF 层。初步步骤是对输入的航行通告文本进行分词处理,随后将这些分词转化为向量,作为模型的输入数据。接着,利用 BiLSTM 层来抽取文本的特征信息,并通过 CRF 层来得出最终的识别成果。

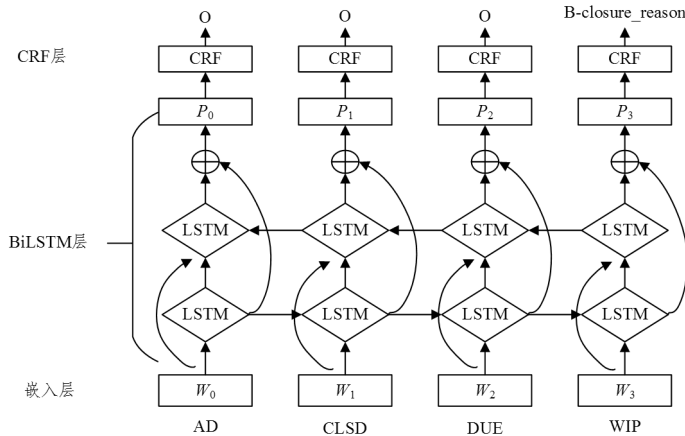


图 2 BiLSTM-CRF 模型结构

Fig. 2 BiLSTM-CRF model structure

#### 3.2 嵌入层

该层主要负责对输入的词进行词向量的映射。首先是对已标注的航行通告文本数据集进行预处理,包括移除冗余的特殊字符。接下来,初始化一个维度为  $n * d$  的矩阵,其中  $n$  代表航行通告文本长度, $d$  表示词向量的维度。这个  $n * d$  维的向量矩阵随后被用作模型输入。

#### 3.3 BiLSTM 层

在文本航行通告命名实体识别任务中,实体识别不仅与过去时刻的信息有关,还与后续的信息有关,即在模型训练中需要访问过去和未来的输入特征,因此采用 BiLSTM 作为深度神经网络模型。BiLSTM 由前后两层 LSTM 组成,前后两层分别按照时间序列前向和时间序列后向,在  $t$  时刻的隐状态分别定义为  $\vec{h}_t$  和  $\overleftarrow{h}_t$ ,其中  $\vec{h}_t$  为前向的 LSTM, $\overleftarrow{h}_t$  为后向的 LSTM,具体计算式如式(1)~式(3)所示:

$$\vec{h}_t = f(U^{(1)} h_{t-1}^1 + W^{(1)} x_t + b^{(1)}) \quad (1)$$

$$\overleftarrow{h}_t = f(U^{(2)} h_{t+1}^2 + W^{(2)} x_t + b^{(2)}) \quad (2)$$

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t \quad (3)$$

其中,  $\oplus$  为向量拼接操作, $h_t$  为双向 LSTM。将词向量序列输入 BiLSTM 层,前向 LSTM 得到词向量的前向隐藏层序列  $\vec{h}_t = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_t\}$ ,后向的 LSTM 则得到词向量的后向隐藏层序列  $\overleftarrow{h}_t = \{\overleftarrow{h}_1, \overleftarrow{h}_2, \dots, \overleftarrow{h}_t\}$ ,按照位置拼接得到最终的隐藏层序列  $h_t = [\vec{h}_t, \overleftarrow{h}_t] \in R^{2 * m}$ ,并作为下一层的输入。

#### 3.4 CRF 层

仅依赖于 BiLSTM 层,可能会导致输出标签序列无效的问题。为解决该问题,提出将 CRF 层连接到 BiLSTM 的输出之后。在序列标注任务中,CRF 模型的作用范围覆盖整个句子结构,而非句中孤立的单个元素。因此,利用 BiLSTM 层产生的输出特征作为 CRF 层的输入,该模型能学习到特定的约束条件,进而执行后续的序列分类和标注工作。

设输入序列为  $x = \{x_1, x_2, \dots, x_n\}$ ,经 BiLSTM 层输出的概率矩阵为  $P_{n * k}$ ,  $P_{i,j}$  表示  $x_i$  被标记为第  $j$  个标签的概率。 $A_{i,j}$  则表示概率转移矩阵中第  $i$  个标签转移到第  $j$  个标签概率。

对于输入序列对应的输出标签序列  $y = \{y_1, y_2, \dots, y_n\}$ ,路径得分公式如下所示,通过计算得分,并输出最佳的标签序列。

$$S(x, y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=0}^n P_{i, y_i} \quad (4)$$

$$y^* = \operatorname{argmax} S(x, y) \quad (5)$$

### 4 数据获取和标注

为验证本文提出的 BiLSTM-CRF 命名实体识别模型在航行通告文本的有效性和可行性,本章使用自建的航行通告语料作为实验数据集。

#### 4.1 数据获取

本文从国际民航组织提供的历史航行通告数据库中提取

了全球有关跑道关闭场景的航行通告,总计 2096 条,将数据按照 8:2 的比例划分为训练集和测试集。其划分的数据集中训练集总计 1676 条,测试集总计 420 条。

#### 4.2 数据标注

本次标注任务采用 doccano 标注工具,doccano 是一个开源的文本标注工具,可以用它为自然语言处理的语料进行标注,操作非常简单,在小型语料库上,只需数小时就能完成全部的标注工作。借助该标注工具对 airport\_designator, runway, runway\_direction, runway\_surface\_composition, closure\_reason, start\_time, end\_time, schedule, note 9 类实体进行标注,完成后生成 json 文件,该文件中存放了所标注实体的位置索引信息;然后基于 Python 编程语言,使用 BIO 实体标注法(B 为实体的开始,I 为实体的内部,O 为非实体)给每个词打上对应的标签(见表 1)。其中实体标注数量统计如表 2 所列。

表 1 数据标注示例

Table 1 Example of data annotation

词序列	标注序列
RWY	O
16R/34L	B-runway
CLSD	O
DUE	O
TO	O
WIP	B-closure_reason
.	O

表 2 标注数据统计

Table 2 Statistics of labeled data

标注实体	实体含义	标注数量
airport_designator	机场代码	2094
runway	跑道	2014
runway_direction	跑道方向	74
runway_surface_composition	跑道道面类型	118
closure_reason	关闭原因	1180
start_time	通告生效时间	2096
end_time	通告失效时间	2096
schedule	时间表	876
note	注释	840

## 5 实验分析

### 5.1 实验环境及参数设置

本次实验采用 Pytorch2.1.2 深度学习框架和 Python3.10 环境,硬件配置为 6 核 Intel Core i7 处理器,32 GB 内存,Radeon Pro 560x 4 GB 显卡。训练过程中,采用 Adam 优化器,学习率设置为 0.001,batch\_size 为 16,为了防止过拟合问题,对 BiLSTM 的输入和输出中使用 Dropout,取值为 0.5,具体训练参数设置如表 3 所列。

表 3 参数配置表

Table 3 Parameter configurations

参数类别	取值
embedding_dim	100
hidden_dim	256
batch_size	16
学习率	0.001
优化器	Adam
dropout	0.5
epoch	50

### 5.2 模型评价指标

本文涉及的命名实体识别任务有九类,需对单个类别和

整体识别的性能进行评估。实验采用最常见的准确率(Precision,  $P$ )、召回率(Recall,  $R$ )和调和均值( $F1$ -score)作为评价指标对模型的最终性能进行评估,具体的评估公式如下:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (6)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

$$F1 = \frac{2PR}{P + R} \times 100\% \quad (8)$$

其中, $TP$  表示识别出正确实体的个数, $FP$  表示识别出不正确实体的个数, $FN$  表示未识别出实体的个数。由上述公式可知,如果希望创建一个具有最佳准确率-召回率平衡的模型,则需要尝试将  $F1$  值最大化。

### 5.3 模型实验结果分析

本次实验采用 LSTM, BiLSTM, BiLSTM-CRF 3 种模型在相同训练集和测试集上进行 50 次迭代对比实验。表 4 列出了 BiLSTM-CRF 模型算法对 9 类实体识别的结果。从表 2 中可以看出,本文采用的 BiLSTM-CRF 模型的整体识别效果非常显著,尤其是在一些特征比较明显的实体上,如 airport\_designator, start\_time, end\_time,  $F1$  值为 100%。对于有些实体识别效果低于 90% 的原因主要为,当前 ICAO 航行通告数据库中含有大量的历史航行通告数据,这类历史数据常包含不规范用语,行业内称为“不规范航行通告”,因而对于这些数据,模型整体上并不能够较好地识别,尤其是对于 note 实体,该数据段中常包含特殊符号加文字的专业术语组合描述,语义结构比较复杂,模型难以学习到该类实体的特征信息,效果最差, $F1$  值仅为 66%。

表 4 BiLSTM-CRF 模型算法结果

Table 4 Algorithm results of BiLSTM-CRF model

实体类别/整体	$P$	$R$	$F_1$
airport_designator	100	100	100
closure_reason	85	90	88
end_time	100	100	100
note	66	65	66
runway	99	99	99
runway_direction	100	86	92
runway_surface_composition	81	100	90
schedule	98	98	98
start_time	100	100	100

表 5 列出了各组模型的整体实验结果。从表中可知,对 LSTM 和 BiLSTM 结果的分析,证明了双向的 LSTM 确实能够通过上下文信息提取到更加丰富的特征,其效果相较于 LSTM 提升了 5 个百分点,而对于在加入 CRF 层的 BiLSTM,同样其  $F1$  值高于未加入 CRF 层的 BiLSTM,效果提升 3 个百分点,验证了 CRF 层的引入显著提升了模型的性能,这是因为 CRF 通过学习标签间转移概率的矩阵实现的,加强了标签之间的关联性,进而使模型能够更加充分地利用序列的整体信息。

表 5 3 种模型实验结果

Table 5 Experimental results of three models

模型	$P$	$R$	$F_1$
LSTM	82	92	87
BiLSTM	92	93	92
BiLSTM-CRF	<b>95</b>	<b>95</b>	<b>95</b>

最后从自建的航行通告语料库中随机抽取文本航行通告对模型进行测试,结果如表 6 所列。

测试结果验证了本文提出的 BiLSTM-CRF 模型可以有

效从航行通告文本中抽取所需的实体,为后续研究工作实现文本航行通告转换为数字航行通告数据格式提供了技术实现基础。

表 6 BiLSTM-CRF 模型实体识别效果示例  
Table 6 Example of BiLSTM-CRF model entity recognition effect

航行通告文本	实体识别结果
K4024/23 NOTAMR K3016/23 Q ) CZUL/QMRLC/IV/NBO/A/000/999/5013N06616W005 A ) CYZV B ) 2312301144 C ) 2403272000 E ) RWY 13/31 CLSD.	airport_designator:CYZV start_time:2312301144 end_time:2403272000 runway:13/31
E0283/24 NOTAMR E7091/23 Q ) CZUL/QMRLC/IV/NBO/A/000/999/4541N07402W005 A ) CYMX B ) 2401161248 C ) 2404161200 E ) RWY 11/29 CLSD DUE NO MAINT.	airport_designator:CYMX start_time:2401161248 end_time:2404161200 runway:11/29 closure_reason:NO MAINT
M0668/24 NOTAMR M5716/23 Q ) EBBU/QMRLC/IV/NBO/A/000/999/5015N00439E005 A ) EBFS B ) 2402070713 C ) 2405060800 EST E ) RWY 08R/26L CLSD DUE TO WIP	airport_designator:EBBU start_time:2402070713 end_time:2405060800 EST runway:08R/26L closure_reason:WIP

## 6 航行通告命名实体识别系统

### 6.1 系统设计

航行通告命名实体识别系统的设计如图 3 所示。

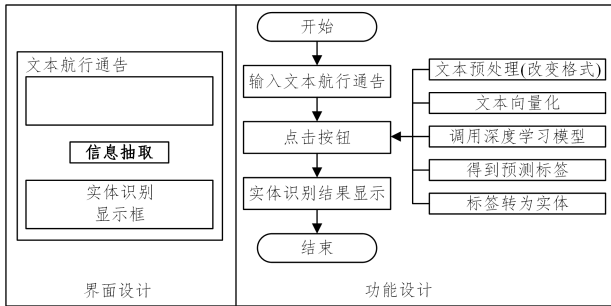


图 3 系统设计  
Fig. 3 System design

### 6.2 系统实现

系统界面及功能基于 Python 编程语言实现,其中界面通过 Tkinter 实现。对于该命名实体识别系统,首先运行系统,进入到系统界面,在界面中包含有文本航行通告输入框、信息抽取按钮以及实体识别显示框 3 部分,具体操作步骤为:首先在文本航行通告输入框中输入所要识别抽取的文本航行通告,其次点击信息抽取按钮,有关该文本航行通告中的所有命名实体被识别抽取出来并显示在实体显示框中,输出界面如图 4 所示。



图 4 系统界面输出结果

Fig. 4 System interface output results

**结束语** 本文首次将深度学习应用到航空情报领域中航行通告部分,针对航行通告自建的数据集进行人工标注,设计并训练 BiLSTM-CRF 模型。在实验中模型取得了 95% 的精确率、95% 的召回率以及 95% 的 F1 值,实体识别效果明显优于 LSTM 和 BiLSTM 两种模型。并设计了简易的航行通告命名实体识别系统,利用训练的模型实现了信息识别抽取的功能,为后续航行通告系统研发中将文本航行通告转换为数字航行通告数据格式提供了技术支持。实现文本航行通告转换为数字航行通告的格式,将有效改善文本航行通告在数字环境的兼容问题。尽管当前本文提出的模型在航行通告实体识别上取得了不错的效果,但在未来的研究中仍有改进的空间。后续工作可以从以下几个方向考虑:

- 1) 当前的词嵌入部分通过随机生成的数值分配给对应的词,该类方式的词向量缺失语义之间的关联。下一步将通过收集更多的航行通告数据,采用 word2vec 模型训练生成更加丰富的词向量,或考虑采用 BERT 模型作为嵌入层。
- 2) 航行通告涉及多场景(跑道关闭、机场关闭、空域限制、导航设施不可用等等),为方便构建模型,本文当前仅考虑单一场景,使得数据集规模较小,后续将考虑更多的航行通告场景,增大数据集规模,利用深度学习模型可先基于场景对航行通告进行分类,之后再具体场景的命名实体识别。

### 参考文献

[1] CAAC. Specification for the Preparation and Issuance of Civil Aviation NOTAMS [Z]. 2011;30P. ;A4.

[2] EUROCONTROL. Digital NOTAM Specification [EB/OL]. [2024-03-15]. [https://ext.eurocontrol.int/aixm\\_confluence/display/DNOTAM/Digital+NOTAM+Specification](https://ext.eurocontrol.int/aixm_confluence/display/DNOTAM/Digital+NOTAM+Specification).

[3] GUO X Y, HE T T. Survey about Research on Information Extraction [J]. Computer Science, 2015, 42(2): 14-17, 38.

[4] GAO X, WANG S, ZHU J W, et al. Overview of Named Entity Recognition Tasks [J]. Computer Science, 2023, 50(S1): 26-33.

[5] RABINER L R. A tutorial on hidden Markov models and selected applications in speech recognition [J]. Proc IEEE, 1989, 77.

[6] PIETRA S D, PIETRA V D, MERCER R L, et al. Adaptive language modeling using minimum discriminant estimation [C] //

- IEEE International Conference on Acoustics. IEEE, 1992; 633-636.
- [7] LAFFERTY J, MCCALLUM A, PEREIRA F C N. Conditional Random Fields; Probabilistic Models for Segmenting and Labeling Sequence Data [C] // Proceedings of the 18th International Conference on Machine Learning (ICML 2001). 2001; 282-289.
- [8] CHEN P H, LIN C J, SCHLKOPF B. A Tutorial on v-support vector machines [J]. Applied Stochastic Models in Business and Industry, 2005, 21(2): 111-136.
- [9] BIKEL D M, SCHWARTZ R, WEISCHEDEL R M. An Algorithm that Learns What's in a Name [J]. Machine Learning, 1999, 34: 211-231.
- [10] BORTHWICK A, STERLING J, AGICHTEN E, et al. Description of the MENE named entity system as used in MUC-7 [C] // Proceedings of the 7th Message Understanding Conference, Virginia. Stroudsburg: ACL, 1998; 1-7.
- [11] MCCALLUM A, LI W. Early Results for Named Entity Recognition with Conditional Random Fields, Feature Induction and Web-Enhanced Lexicons [J]. Association for Computational Linguistics, 2003, 4: 188-191.
- [12] TAKEUCHI K, COLLIER N. Use of Support vector machines in extended named entity recognition [C] // Proceedings of the 6th Conference on Natural Language Learning, Taipei, China, Aug 24—Sep 1, 2002. Stroudsburg: ACL, 2002; 184-190.
- [13] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural Language Processing (almost) from Scratch [J]. Journal of Machine Learning Research, 2011, 12(1): 2493-2537.
- [14] HUANG Z, XU W, YU K. Bidirectional LSTM-CRF Models for Sequence Tagging [J]. arXiv: 1508. 01991, 2015.
- [15] CETOLI A, BRAGAGLIA S, O' HARNEY A D, et al. Graph convolutional networks for named entity recognition [C] // Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories. Stroudsburg: ACL, 2018; 37-45.
- [16] DEVLIN J, CHANG M W, LEE K, et al. BERT; Pre-training of Deep Bidirectional Transformers for Language Understanding [C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019; 4171-4186.
- [17] GUO Z X, DENG X L. Intelligent Identification Method of Legal Case Entity Based on BERT-BiLSTM-CRF [J]. Journal of Beijing University of Posts and Telecommunications, 2021, 44(4): 129-134.
- [18] LIU H, ZHANG Z X, WANG Y F. A Review on Main Optimization Methods of BERT [J]. Data Analysis and Knowledge Discovery, 2021, 5(1): 3-15.
- [19] LI H F, ZENG C, HU H Q, et al. Research on Text Classification of NOTAM Based on Machine Learning [J]. Journal of Civil Aviation, 2022, 6(4): 6-9.
- [20] XIANG H, ZHANG C, LI M. Identification method for non-normative NOTAM based on NLP [J]. Journal of Civil Aviation University of China, 2022, 40(2): 14-18.
- [21] PAN Z X, LUO Y H, LI R Z. Research on Extraction of Notam Information [J]. Modern Computer, 2022, 28(2): 82-87.



**XIANG Heng**, born in 1975, master, associate professor. His main research interests include air traffic control human factors and aeronautical information management.