**引用格式**：郭轩，彭宏，魏莱.语义融合的革命文物图像以文标图算法研究[J].中国传媒大学学报（自然科学版），2022，29(04)：26-32+49.
**文章编号**：1673-4793(2022)04-0026-08

# 语义融合的革命文物图像以文标图算法研究

郭轩[1]，彭宏[2*]，魏莱[1]
（1.北京邮电大学计算机学院，北京 100876；2.文化和旅游部民族民间文艺发展中心，北京 100007）

**摘要**：革命文物蕴含着丰富的红色文化和光荣历史，具有重要的研究价值和传承意义。但目前对革命文物的梳理和解读仍缺乏数字化的方法。基于革命文物的多模态数据组织形式，本文提出了一种全新的面向革命文物图文数据的领域化、标签化和结构化方法——“以文标图”，对革命文物进行数字化标注。针对革命文物图像标签化问题，构建多模态语义融合模型提取图像标签，使用多特征TF-IWF方法提取文本标签，最后基于标签语义相似度对标签重排序，得到图像相关性高、信息粒度细的图像标签。针对革命文物图文数据结构化，构建图文模态融合模型帮助数据结构化，并将传统的基于序列标注的命名实体识别方法转化为属性名预测和属性值预测两部分。本文算法实现了图文模态的语义信息互补，提高了图文数据标签化和结构化的效果，为革命文物信息标注和解读提供了技术路径。

**关键词**：革命文物；多模态语义融合；关键词提取；命名实体识别
**中图分类号**：TP391 **文献标识码**：A

# Image-text annotation for revolutionary cultural relics based on semantic fusion

GUO Xuan[1], PENG Hong[2*], WEI Lai[1]
*(1.School of Computer, Beijing University of Post and Telecommunication, Beijing 100876, China; 2.Center for Ethnic and National Folk Literature and Art Development，Ministry of Culture and Tourism, Beijing 100007, China)*

**Abstract:** Revolutionary cultural relics contain rich red culture and glorious history, and have important research value and inheritance significance. However, there is still a lack of digital methods for the sorting and interpretation of revolutionary cultural relics. Based on the multimodal data form of organization of revolutionary cultural relics, we propose a innovative method of territorialization, labeling and structuring for revolutionary cultural relic image --"image-text annotation", which digitally label revolutionary cultural relics. For the problem of image tagging of revolutionary cultural relics, a multi-modal semantic fusion model is constructed to extract image tags, and the multi feature TF-IWF method is used to extract text tags. Finally, the tags are reordered based on the semantic similarity of them, and the tags with high correlation of image and fine granularity of information are obtained. For structuring of the image and text data of revolutionary cultural relics, a image modal and text modal fusion model is constructed. The traditional named entity recognition method based on sequence annotation is transformed into two parts: attribute name prediction and attribute value prediction. The algorithm in this paper realizes the complementation of semantic information between image modal and

**基金项目**：揭榜挂帅重点研发课题（课题编号：2021YFF0901701）
**作者简介（*为通讯作者）**：彭宏（1972-），男，高级工程师，主要研究领域为文化资源数字化。Email：466985365@qq.com

text modal, improves the result of labeling and structuring of the image and text data, and provides a technical path for the annotation and interpretation of revolutionary cultural relics information.

**Keywords:** revolutionary cultural relics; multimodal semantic fusion; keyword extraction; named entity recognition

## 1　引言

革命文物记载了共产党人一路走来的伟大历程和感人事迹，标注并解读革命文物背后的信息，能够使人民感悟红色基因红色文化，有利于加强社会主义精神文明建设，提高爱国主义热情。2022年1月，财政部、国家文物局联合颁布了《关于加强新时代革命文物工作的通知》[1]，通知指出要加强革命文物数字化保护和数字化传播。在馆藏文物的基础上，引入革命文物数字化采集、数字化存储、数字化标注，建立在线红色基因库。建设红色基因库不仅能够服务于红色文化检索、藏品管理、知识图谱、文物溯源等下游应用，也能够实现革命文物信息自动语义理解，是建设“三库”（标本库、基因库、素材库），消弭“语义鸿沟”的重要手段。本文对革命文物中的多模态信息进行标签化、结构化，最终实现文物数字化标注，是构建红色基因库的重要步骤。

革命文物数据一般来源于博物馆实物、领域相关图书、网络资源，通常包含一组文物图像和文物介绍文本，属于强相关图文信息。如表1展示了“红军路过扎西时留下的手榴弹”对应的文物数据。因而对革命文物数据标签化和结构化时应同时借助图像特征和文本特征。本文首先构建了革命文物标签体系，包括图像分类标签和文本属性标签。为了得到专业性更强、粒度更细的图像标签，本文使用改进的多特征TF-IWF（Term Frequency-Inverse Word Frequency）提取文本关键词作为文本标签，使用图像和文本融合特征得到图像标签。通过基于混合策略的词义相似度算法对文本标签的重排序，选取与图像标签相似度高的文本标签作为图像标注结果。上述图像标签化方法兼顾标签的领域专业性与图片的相关性，提升了图像标签的质量。为了充分利用图文特征进行革命文物数据结构化，本文利用改进的Co-Transformer模块对图文模态特征进行语义融合，并将传统的基于序列标注的命名实体识别方法转化为属性名预测和属性值预测两个子任务。本文的数据结构化方法充分利用图文两个模态的特征，提高了革命文物数据结构化的质量。

**表1“红军路过扎西时留下的手榴弹”对应的文物数据**

| 图像 | 标题 | 描述文字 |
| --- | --- | --- |
|  | 红军路过扎西时留下的手榴弹 | 年代：民国 这三枚红军手雷，就是红军在川滇黔边区进行游击和机动作战时留下的。当时红军撤离时，将手雷埋入土中。1958年大跃进期间，才在土中挖出了这些手雷。这些手雷都有明显的五角星，五角星内是象征共产党的镰刀锤子。 |

## 2　相关工作

本文工作集中在面向图文数据的多模态内容标注上。目前，面向图文数据的多模态内容标注的实现方式主要有转化为图文多标签分类问题、转化为图文跨模态检索问题、转化为图像描述生成问题。

图文多标签分类与多类分类不同，是指将一个图文样本标注为一或多个标签。在面向图文数据的多模态内容标注中，通常先对图文特征进行融合，然后使用融合特征进行多标签分类，或者针对不同模态构建多个多标签分类器，然后通过决策手段整合结果标签。例如Oramas[2]等在包含250个类别的MuMu音乐多模态数据集中，融合了音频、文本、图像三模态的特征对3万张专辑进行了多标签分类。Chen[3]等联合地训练深度多模态网络与多标签分类器，借模态和标签的关系提高了多标签分类性能。在Transformer结构

应用于多模态领域后,借助迁移学习的思想,将预训练模型迁移到图文多模态数据上做微调,并对接多标签分类作为下游任务,能以最小的数据训练量完成内容标注需求。

图文跨模态检索分为以图搜文(Image Annotation)和以文搜图(Image Search),以文搜图指输入一段文本检索与文本最相关的图像簇,以图搜文指输入一副图像检索与图像最相关的文本簇。图文跨模态检索需要考虑异质数据的相似度度量问题。在面向图文数据的多模态内容标注中,首先对文本进行分词和编码,将编码结果输入到以图搜文模型进行单词与全局图像或局部图像的相关性对齐,然后按照图文相关性高低对候选标签排序,选择相关性较高的候选标签作为标注结果。Srivastava[4]等通过深度玻尔兹曼机(Deep Boltzmann Machines)生成图文联合空间概率分布,对于输入的图文,根据条件概率P(图像|文字)或P(文字|图像)在联合空间中检索出相关图像或文本,将相关文本作为标签。基于跨模态哈希的图文检索[5],将图文数据编码为哈希值映射到汉明空间,在汉明空间中进行高效率的图文标注。

图像描述生成(Image Caption)任务是根据输入图像特征输出一段与图像相关的描述文本,早期的图像描述生成常常构建固定句子模版,预留若干单词槽对应图像中的目标检测框,然后对检测框分类,最后将分类标签赋值到单词槽中完成描述文本,例如Baby Talk[6],但此类方法受限于句子模版,于是2015年后开始使用机器翻译领域常用的编码器-解码器的结构,类似于语言翻译,把图像视作待翻译语言,使用CNN、VGG等提取图像特征作为编码,使用RNN或LSTM作为解码器输出每个单词对应的概率,从而生成描述文本。

除了上面三种主流方法外,多模态内容标注的工作还有面向传统服饰领域的以文标图[7]、VQA模型的方法[8][9]、结构化信息抽取的方法[10][11]等,大多基于深度学习模型完成对多模态信息的融合与理解。

## 3 革命文物标签体系构建

革命文物图像分类标签主要关注革命文物图像的分类,分类标签会应用在第4节基于多模态语义特征融合的图片标签提取任务中,作为多标签分类的依据。革命文物图像分类标签体系学习了多地革命博物馆的分类思路,从革命文物图像特征出发,综合考虑了文物的材质、用途,制定了以下的两套分类标准。按材质分类为:玻璃制品、皮革制品、金属制品、塑料制品、纤维织物、石制品、竹木制品、宝玉石制品、陶瓷制品、其他材质。按用途分类为:图书宣传品、服装饰品、档案文书、家具炊具、书法绘画、奖章印章、货币票据、工艺品、武器装备、容器、交通运输、文具、机械装置、其他杂类。

革命文物结构化标签体系主要关注革命文物文本中词语的属性归类,属性类别会被应用于第5节中图文内容结构化算法,作为序列标注的依据。革命文物结构化标签体系参考了革命文物相关元数据标准,最终确定了15种属性:时间、地点、组织、事件、人物、颜色、材质、形状、书名、数字、型号、货币、重量、长度、时长。

## 4 基于多模态语义融合的革命文物图像标签化

革命文物数据往往包含文物图像和周围文本描述信息,单纯地使用图像自动标注算法对图像进行标注往往会丢失文字模态的语义信息。其次,目前的图像自动标注算法标注的结果仅仅局限于分类标签列表,难以实现对图像内容的细粒度描述。最后,革命文物具备历史文化领域属性,没有领域专家的参与很难得到专业可信的标签。因此,本文运用多模态语义融合的多标签图像分类模型提取图像标签,运用改进的TF-IWF关键词提取模型提取文字标签,将图像标签和文字标签作为标注结果。之后,通过基于混合策略的词义相似度算法对标注结果的重排序,提升了图文语义相关标签的置信度。

### 4.1 基于多模态语义特征融合的图像标签提取

为了使用文本语义信息,提升图像标签的准确性和专业性,本节设计了基于图文特征语义融合技术的图文多标签分类网络模型,结合第3节构建的图像标签列表对革命文物图文的数据进行了多标签分类,分类结果作为图片的标签。

根据中文语料的特点,将文本中的字嵌入为低维向量,将字向量拼接形成句子向量作为文本编码。然后在文本编码前后分别添加特殊符号([CLS]和[SEP])用于分类和分隔不同片段。对于图像,使用ResNet-101预训练模型提取平均池化层前的局部特征图作为图像编码。另外,数据集中往往会出现图像缺失或者文本缺失的情况。为了尽可能避免模态缺失导致的语义信息损失,算法的输入向量添加了第三个片段,第三个片段由文本片段和图像片段加权平均

得到。将三个片段拼接输入到Bert中提取多模态语义融合特征。多标签分类任务可以转化为判断每个标签是否属于某个实例的二分类任务，最后对融合后的特征向量通过一层全连接层进行二分类。模型结构如图1所示。

给定一个标签集I，我们判断其中第$i$个标签是否属于某个实例。训练模型用到的损失函数为二分类交叉熵，公式如下：

$$\text{Loss} = -(y \cdot \log(\hat{y}) + (1-y) \cdot \log(1-\hat{y})) \quad (1)$$

其中y是对$i$标签的二分类结果，1是正例，代表属于该标签，0是负例，代表不属于该标签。$\hat{y}$是对第$i$个标签的预测是正例的概率。

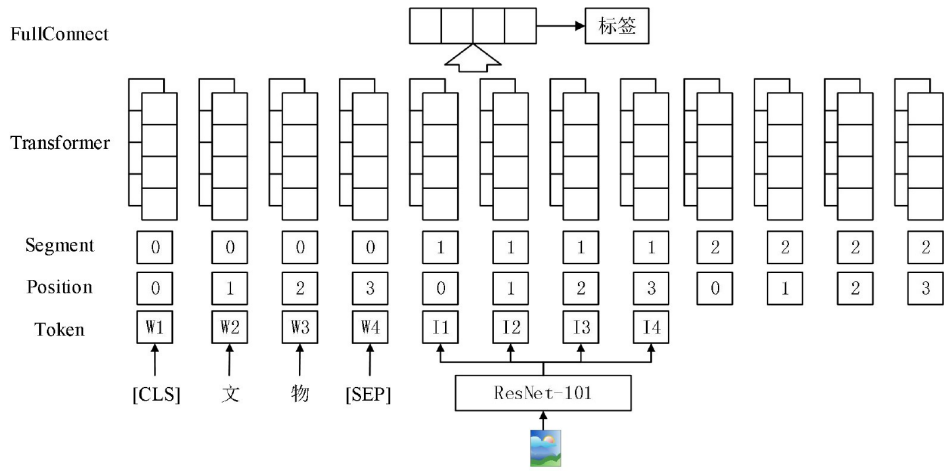


**图1 革命文物数据文物信息标签化、结构化流程**

### 4.2 基于多特征改进的TF-IWF的文本标签提取

文本标签提取有很多方法，例如基于词频特征的TF-IDF，基于词共现模型的TextRank，基于主题模型的LDA等。其中TextRank和LDA对领域短文本的挖掘效果经实验验证不如基于词频特征的方法。TF-IDF是一种基于词频统计的并通过逆文档频率加权抑制噪声的文本标签提取方法，在单一领域语料库中会错误地抑制高频领域标签。例如革命文物领域语料库中“中华人民共和国”会出现在多数文档中，导致逆文档频率很高，相应的权重会被削弱。除此之外，TF-IDF还忽略了词位置、词长度、词性等信息导致预测的语义性进一步下降，因此本文基于多特征改进的TF-IWF结合第3节构建的领域词库对文本进行标签提取。

TF-IWF是对TF-IDF的一种改进，将逆文档频率替换为逆词频，逆词频（IWF）是指全局语料库切分出的所有词语总数与全局语料库中该词的数目的比值，可以抑制超高频的无意义词语如“因为”、“然后”等，但避免了TF-IDF在单一领域语料库中对高频领域词汇的抑制。另外，词位置和词长度能提供重要信息来帮助标签提取：（1）出现在标题中的词比出现在文本中的词更有代表性，应该分配更高的权重。（2）长度越长的单词信息量越大，描述越具体，更有可能成为文本标签。最终选取多特征加权TF-IWF值较高的单词作为文本标签。

给定革命文物文本集合$G$，其中第$j$个文本的第i个单词的多特征加权TF-IWF计算公式如下：

$$TF-IWF_i = \text{Pos}_i \times Len_i \times IWF_i \times TF_{i,j} \quad (2)$$

其中$Pos_i$为单词的位置特征权重，$Len_i$为单词的长度特征权重。该单词的词长度权重计算公式如下：

$$\text{Len}_i = \frac{\text{length}(i) - u}{\partial} \quad (3)$$

其中$u$代表文档中平均词长度，$\partial$代表文档中词长度的方差。该单词的TF-IWF计算公式如下：

$$TF_{i,j} \times IWF_i = \frac{N_{i,j}}{\sum_k N_{k,j}} \times \log \frac{\sum_{i=1}^{m} N_{i,G}}{N_{i,G}} \quad (4)$$

其中$k$代表第$j$个文本对应的单词集合中的单词，$N_{i,j}$代表第$j$个文本第i个单词的数量，$m$代表全局语料库中不重复单词的数量，$N_{i,G}$代表全局语料库中第i个单词的数量。

### 4.3 基于图文标签语义相关性的标签重排序

在分别提取了图片和文本的标签后,如果直接组合作为标注结果,一方面,结果无法表达出标签与图像、标签与文本间的相关性,很可能引入无关标签。另一方面,根据革命文物数据的特点,在多模态多标签分类模型中得到的图像标签是对文物图像特征的概括,信息粒度较粗,但标签与图像高度相关;而由TF-IWF提取的相关文本标签则粒度较细,但标签不一定与图片相关。综上所述,本文从革命文物数据的标签特点出发,提出了基于图文标签语义相关性的标签重排序方法,实现图像标签和文本标签的优势互补,输出图文相关性高、信息粒度细的标注结果。

本文提出的基于图文标签语义相关性的标签重排序算法通过词义相似度计算,以文本标签为主标签,分别计算图像标签和文本标签之间的词义相似度。词义相似度高的标签更可能是图文相关性高的标签,根据词义相似度结果,对文本标签集合从高到低进行排序,相似度高于阈值的标签作为图文对标签,低于阈值的标签可以作为次级标签,作为图文对标签的补充。分别将图文标签表示为词向量,利用余弦相似度计算文本标签的置信度:

$$\mathrm{sim}\ W2V\left(w_1,w_2\right)=\frac{w_1\cdot w_2}{\left\|w_1\right\|\left\|w_2\right\|} \tag{5}$$

其中$w_1$、$w_2$代表两个词向量。

## 5 多模态命名实体识别的革命文物图文数据结构化

革命文物图文数据结构化用于提取图文内容的结构化实体,是以文标图算法的第二部分。革命文物图文数据结构化是对革命文物数据的组成、特征、属性等信息进行结构化描述的过程,通过图文数据结构化能够抽取革命文物的元数据信息。本文基于Transformer模块设计了一种多模态命名实体识别模型,将提取结构化实体的过程分为两部分:属性名预测和属性值预测。属性名指革命文物结构化标签体系的15种属性,属性值指句子中被标注为某个属性的单词或字的组合。模型的具体结构如图2所示。

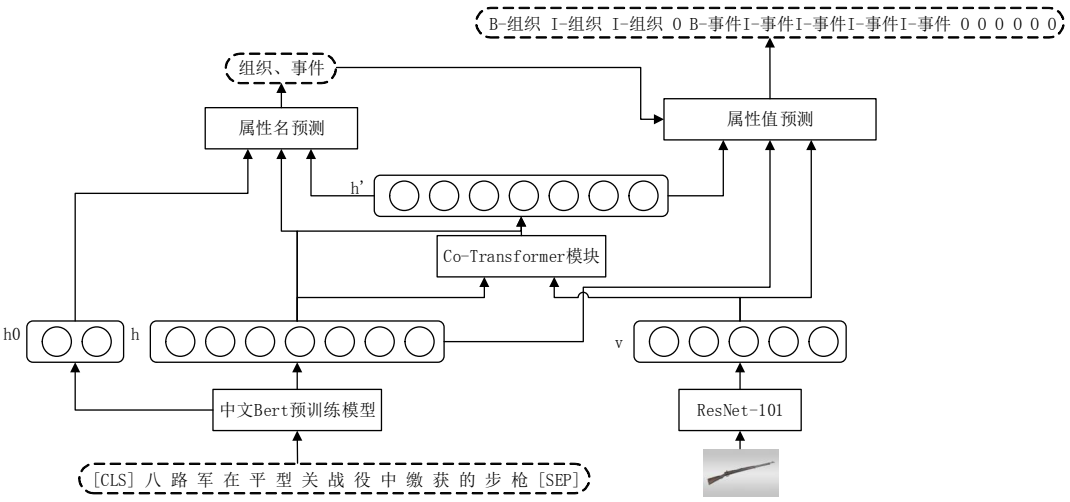


图2 多模态命名实体识别模型结构图

### 5.1 图文编码器

对于文本,模型使用了Bert中文预训练模型对文字进行特征编码,该模型采用字向量(Token Embedding)、位置向量(Positional Embedding)、片段向量(Segment Embedding)三者加和后的向量作为特征向量。对于图像,采用了ResNet-101预训练模型的第五层卷积层的输出结果进行维度堆叠后作为图像视觉特征$v$:

$$v=\left(v_1,\ldots,v_k\right) \tag{6}$$

其中,$v_k$表示局部特征图的向量表示。ResNet-101已经在ImageNet数据集上预先训练完成,能够对通用领域物体提取较为精准的语义特征。

### 5.2 图文模态融合模块

本文设计的模型使用基于Transformer结构改进的Co-Transformer模块对图文模态特征进行语义融合,通

过Co-Transformer模块能够在图像的帮助下有选择地丰富文本特征的语义表达，在文字的帮助下有选择地丰富视觉特征的语义表达。Transformer 编码器由多个标准单元串联组成。其中第l+1层单元接受上一层所提取的特征，特征向量会分别乘上三个不同的可训练的权重矩阵进行线性变换，得到 Query、Key、Value 三种向量后输入到自监督注意力模块中。在Co-Transformer 结构中，视觉模块和和文本模块分别由多个标准单元串联组成。在表征考虑文本上下文的视觉特征时，视觉模块的第i+1层单元接受第i层的图像特征和第j层的文本特征；输入的 Query 来自于图像特征，Key 和 Value 来自于文本特征。在表征考虑图像相关区域的文本特征时，文本模块的第j+1层单元接受第i层的图像特征和第j层的文本特征；输入的 Query 来自于文本特征，Key 和 Value 来自于图像特征。Transformer 和 Co-Transformer 结构如下图 3 所示：

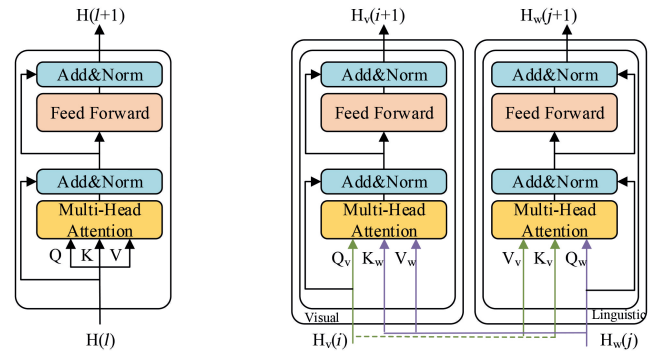


**图3 Co-Transformer 结构图**

### 5.3 属性名预测

根据革命文物结构化标签体系的15种属性，使用BIO标注法对训练集语料进行标注，构造以15种属性组成的属性名词袋向量$y^a$，公式如下所示：

$$y^a = \left(y_1^a,\ldots,y_L^a\right) \tag{7}$$

其中$L$为15。如果$y_i^a$为1代表训练集文本的属性名包含属性名词袋中第i个属性名标签。对于每个图文对数据，将 Bert 编码后的文本特征$h_i$、Co-Transformer模块提取的多模态注意力特征$h_i'$以及 Bert 中的特殊分隔符[CLS]的特征$h_0$三者输入到前馈神经网络拟合线性函数，最后经过 Sigmoid 激活函数激活得到预测的属性名词袋向量$\hat{y}^a$。公式如下所示：

$$\hat{y}^a = \sigma\left(W_3\sum_i h_i + W_4\sum_i h_i' + W_5 h_0\right) = \left(\hat{y}_1^a,\ldots,\hat{y}_L^a\right) \tag{8}$$

其中$W_3$、$W_4$、$W_5$为权重矩阵。

### 5.4 属性值预测

属性值的预测可以转化为序列标注的任务，输入的观测序列为单词序列$x = \left(x_1,\ldots,x_N\right)$，输出的属性标记序列为BIO形式的标注序列$y^v = \left(y_1^v,\ldots,y_N^v\right)$。个别属性值的预测与属性名和局部图像具有一定相关性，例如："银"可以是"材质"、"颜色"、"货币"等属性名的属性值，"银"也能够在训练过程中与大量类似的图像建立联系，所以除了使用文本特征和多模态特征外，模型还引入了图像视觉特征和属性名预测标签辅助属性值的预测，使得属性值的预测更为精准。具体地，对于每个图文对数据，将Bert编码后的文本特征$h_i$、Co-Transformer模块提取的多模态注意力特征$h_i'$、图像视觉特征$v_k$以及属性名预测的属性名词袋向量$\hat{y}^a$全部输入到前馈神经网络进行线性拟合，最后经过Softmax激活函数激活得到预测的属性值序列$\hat{y}^v = \left(\hat{y}_1^v,\ldots,\hat{y}_N^v\right)$，其中第i个属性值标签$\hat{y}_i^v$如公式所示：

$$\hat{y}_i^v = \mathrm{softmax}\left(W_6 h_i + W_7 h_i' + W_8\hat{y}^a + \sum_k \alpha_{ik}^v W_V^v v_k\right) \tag{9}$$

其中$W_6$、$W_7$、$W_8$、$W_V^v$为权重矩阵。

## 6 实现及结果

实验数据集为人工整理搜集的4637组革命文物相关图文数据，每组数据包含一张文物图片、图像标题文本、描述文本以及该组数据对应的标签。数据集中的标签包含图像标签和文本标签，以文字标题为"红军路过扎西时留下的手榴弹"的文物为例，文物详细数据如表1所示。该文物对应的图像标签被标注为"金属制品"、"武器装备"，文本标签被标注为"手榴弹"、"红军"、"扎

西”。数据集中的图像标签主要用于训练多模态多标签分类模型,文本标签则作为实验的评价对象的一部分。

本文取精确率、召回率和F-1 Score作为评价标准。由于数据集中正确标签数量和算法标注出的标签数量不一定相等,所以精确率被定义为算法标注出的正确标签数量与算法标注出的所有标签总数的比值,召回率被定义为算法标注出的正确标签数量与数据集中正确标签数量的比值。

### 6.1 革命文物图像标签化实验结果

以文标图标签化基于文本标签抽取,所以实验对比了TextRank、TF-IDF、KeyBert三种文本标签提取算法,取前5个关键词和前10个关键词分别计算准确率、召回率、F-1 Score。实验结果如表2所示:

**表2 文本标签提取实验对比结果**

| 算法 | Top-5 | | | Top-10 | | |
|---|---|---|---|---|---|---|
| | 准确率 | 召回率 | F-1值 | 准确率 | 召回率 | F-1值 |
| TextRank | 0.303 | 0.318 | 0.310 | 0.295 | 0.312 | 0.303 |
| TF-IDF | 0.398 | 0.410 | 0.404 | 0.392 | 0.406 | 0.399 |
| KeyBert | 0.376 | 0.357 | 0.366 | 0.365 | 0.361 | 0.363 |
| 本文算法 | 0.421 | 0.435 | 0.428 | 0.396 | 0.439 | 0.416 |

根据实验结果分析得知,TextRank算法因为只关注文章内部词共现信息,忽略了外部文档的影响,导致该算法无法适应包含词语相对较少的短文本标注,部分短文本甚至完全挖掘不出任何标签。而经典算法TF-IDF则因为基于领域词库分词的全局语料IDF词典,正确标注出了约40%的标签。KeyBert算法的准确率和查准率受训练语料影响较大,如果革命文物训练语料增加,KeyBert可能会有更好的表现。本文算法因为采用了多模态图像多标签分类和多特征文本提取以及图文相关标签重排序,在一定程度上考虑了图文的语义相关性,对数据的标注更为精准,特别是在前5个关键词中的准确率非常高。

### 6.2 革命文物图文数据结构化实验结果

以文标图结构化实验阶段分别对比了CRF、双向LSTM+CRF、Bert+CRF、以及多模态信息抽取模型MAE四组命名实体识别相关的算法,实验结果见表3:

**表3 图文数据结构化实验对比结果**

| 算法 | 准确率 | 查准率 | Micro F1 | Macro F1 |
|---|---|---|---|---|
| CRF | 0.735 | 0.726 | 0.730 | 0.719 |
| LSTM+CRF | 0.826 | 0.784 | 0.804 | 0.793 |
| Bert+CRF | 0.841 | 0.793 | 0.816 | 0.815 |
| MAE | 0.839 | 0.804 | 0.821 | 0.818 |
| 本文算法 | 0.850 | 0.821 | 0.835 | 0.831 |

根据实验结果分析得知,经典算法CRF因为采用人工设计的特征,精度与深度学习提取特征的模型有一定差距。Bert和LSTM结合CRF的算法在CRF的基础上精度提升超过10%。MAE利用了图文模态预测属性,但因为简单拼接的融合方式导致图文语义特征不明显,准确度较低。最后是本文算法,基于Co-Transformer的融合特征令分类精度有了一定提升,虽然相比Bert+CRF性能提升不多,但对于部分属性例如“材质”中的“银”与“颜色”中的“银”,引入图像特征后,对应属性类的F-1值有了一定提升,能够有效避免属性歧义现象。

## 7 总结及未来工作

为了响应革命文物在数字化标注方面新需求,本论文着眼于革命文物图文数据。从网络、博物馆、图书等渠道搜集、整理、标注了数千组革命文物图文数据,基于多模态融合、标签相似度计算、关键词提取、命名实体识别等技术设计了以文标图算法对革命文物图文数据分别做了标签化和结构化。标签化是指提取革命文物主题相关的标签,服务于在线文物推荐、知识检索等场景;而结构化是指标注出革命文物文本的实体位置及其类别如人物、地点、时间等,在文物特征、时间、空间等角度上为革命文物学者的研究提供了参考价值。但目前的研究还存在一些不足需要进一步完善。例如领域词库构建无法自主判别、图文模态语义融合不均衡等,需要进一步研究和探索。

**参考文献(References):**

[1] 文物局,财政部. 国家文物局 财政部关于加强新时代革命文物工作的通知[EB/OL].(2022-01-08)http://www.gov.cn/zhengce/zhengceku/2022-01/08/content_5667074.htm.

[2] Oramas S, Nieto O, Barbieri F, et al. Multi-label music genre classification from audio, text, and images using

(下转第49页)