

融合 CNN 与 BLSTM 神经网络的面部表情识别*

王忠民^{1,2} 侯田田¹ 衡霞^{1,2}

(1. 西安邮电大学计算机学院 西安 710121)

(2. 西安邮电大学陕西省网络数据分析与智能处理重点实验室 西安 710121)

摘 要 为了有效处理面部表情识别中特征提取不完整和特征选择偏差较大等问题,研究了一种新的融合多种卷积神经网络的融合算法。首先通过采用三维卷积神经网络(3-dimensional Convolutional Neural Network, 3D CNN)对连续的多张图片中的特征进行提取。然后把特征信息送至双向长短记忆网络(Bidirectional Long Short-Term Memory, BLSTM)递归神经网络得到这些特征在时域上的映射信息,最后通过 softmax 分类器得到分类结果。通过真实的数据集实验,将新的融合算法与其他几种常见方法进行了对比分析,结果表明能够有效提高分类准确度。

关键词 面部表情识别; 3D CNN; BLSTM

中图分类号 TP391.4 **DOI:** 10. 3969/j. issn. 1672-9722. 2020. 12. 032

Facial Expression Recognition Using CNN and BLSTM

WANG Zhongmin^{1,2} GOU TianTian¹ HENG Xia^{1,2}

(1. School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121)

(2. Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an 710121)

Abstract In order to effectively deal with the problem of feature extraction incompleteness and feature selection bias in facial expression recognition, a new fusion algorithm integrating multiple convolutional neural networks is proposed. Firstly, the facial expression features are extracted from continuous multiple images with using 3D CNN (three-dimensional Convolutional Neural Network). Secondly, these features are sent into the BLSTM recursive neural network model to get the mapping information in time domain. Finally, the probability mapping is used to train a softmax classifier and the final classification result is obtained. Several experiments based on real datasets are made to test the effectiveness of new fusion method with some other methods. The result indicates that new fusion method can effectively improve the classification accuracy.

Key Words facial expression recognition, 3D CNN, BLSTM

Class Number TP391.4

1 引言

人类的情感通常通过其面部表情、声音、手势和姿势等信息表现出来。而面部表情蕴含了人类丰富且细微的情感变化,分析和研究人的面部表情^[1-2]对情感识别^[3-5]、人机交互等方面有很大的应用价值。如今,面部表情识别已经引起了广泛的研

究,同时提出了多种解决方案。而提取面部有效特征^[6-9]成为了面部表情识别中的关键步骤,经典的提特征算法包括独立分量分析法^[10](ICA)、主成分分析法^[11](PCA)、局部二值模型^[12](LBP)。虽然 ICA 等特征提取算法在处理高维数据等方面具有良好的特性,但同时也存在特征提取率低、提取不全面,或是依赖硬件,消耗内存大的等问题^[13]。

* 收稿日期:2020年6月13日,修回日期:2020年7月26日

基金项目:国家自然科学基金项目(编号:61373116);陕西省科技统筹创新工程计划项目(编号:2016KTZDGY04-01);陕西省教育厅专项科研计划项目(编号:16JK1706);西安市科技局科技计划项目(编号:2017084CG/RC047(XAYD001));咸阳市重大科技创新专项(编号:2017K01-25-3);西安邮电大学研究生创新创业基金项目(编号: CXJJ2017061)资助。

作者简介:王忠民,男,博士,教授,研究方向:嵌入式系统,智能信息处理。侯田田,女,硕士研究生,研究方向:深度学习,情感计算。衡霞,女,硕士,副教授,研究方向:模式识别,智能信息处理。

近几年来,利用神经网络^[14]来解决图像识别^[15]问题有了很大的突破,它能让机器自主地从样本数据中学习到表示这些面部表情的本质特征^[16-17],而且更具有一般性。利用3D CNN 更能有效地提取有用的特征信息。采用BLSTM进行分类^[18-19],它能够充分地利用数据的上下文信息^[20],在对序列的识别任务中表现突出^[21],综上所述,本文提出一种融合^[22]3D CNN 和 BLSTM 的面部表情识别方法。面部表情的处理过程如图1所示。

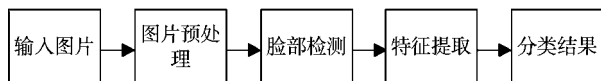


图1 面部表情识别过程图

2 多融合的卷积神经网络

面部表情识别过程主要分为训练和测试两个阶段。在训练阶段,首先对图像进行预处理,然后将预处理之后的图像输入深度卷积神经网络,进而用BLSTM进行分类。在训练阶段,使用已经处理好的训练集通过训练,获得我们所需要的模型;在测试阶段,通过测试集来验证我们模型的性能。如图2所示为面部表情识别系统流程图。



图2 面部表情识别系统流程图

2.1 图像预处理

在输入网络之前,需要对图像进行预处理,使得在之后的模型训练时,得到较为精准的特征,我们采用Adaboost算法对图片中的人脸部位进行精准定位。其核心思想就是针对同一个训练集训练不同的分类器,然后把这弱分类器集合起来,构成一个最终的分类器。在试验过程中,通过迭代

30次主循环,每次循环根据当前的权重分布对样本进行调整,通过学习得到一个弱分类器。而最终的分器是多个弱分类器的加权平均。面部检测过程如图3所示。



图3 面部检测过程

2.2 优化的卷积神经网络模型

本文对经典的卷积神经网络(Convolutional Neural Network, CNN)进行了改进。利用3D CNN 网络在时间维度上的优势和BLSTM在长时间特征依赖关系的优势,提出了新型融合多种卷积神经网络的模型。在该模型种,采用10层结构的卷积神经网络来对面部表情进行识别,包括两个卷积层、两个池化层、4个BLSTM网络层、1个全连接层、1个分类层。如图4所示为面部表情识别流程框架图。

2.2.1 特征提取

在本文中,通过3D CNN 来提取空间和时间维度的特征,从而捕获在多个连续帧中编码的运动信息。3D CNN模型相比于2D CNN,它的性能远远优于2D CNN模型,它能更好地捕获图像序列中的时间和空间的特征信息。因为卷积在3D CNN中具有时空特性,因此它保留了时间信息。

实验中,将连续50张大小为64*64的图片帧作为输入,构建了两个卷积层(C1,C2),两个池化层(S1,S2),1个全连接层(FC1)组成的卷积神经网络结构。卷积核(C1,C2)的大小分别为7*7*3,7*7*3,特征图的大小分别为 $(50-3+1) * (64-7+1) * (64-7+1) = 48 * 58 * 58$ 和 $(16-3+1) * (20-7+1) * (20-7+1) = 14 * 14 * 14$,变量总量分别为 $6 * 48 * 58 * 58 = 968832$ 和 $12 * 14 * 14 * 14 = 32928$ 。池化层的降采样核大小均为3*3,并且全部采用的是最大采样,最终输出的变量规模为 $12 * 5 * 5 * 5 = 1500$ 。每一个池化层都在对应的卷积层之后。如表1所示为网络各层的信息描述。

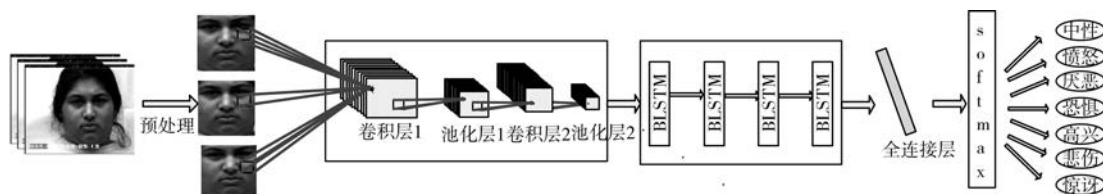


图4 面部表情识别框架

表1 网络各层的信息描述

表示符号	网络层类型	核大小	步长	输出大小
x	输入层			64*64
h_1	卷积层	7*7*3	1	58*58*6
h_2	池化层	3*3*3	2	20*20*6
h_3	卷积层	7*7*3	1	14*14*12
h_4	池化层	3*3*3	2	5*5*12
h_5	全连接层			256
o	输出层			10

2.2.2 分类技术

h_t 长短时记忆 (LSTM) 模型解决了传统 RNN 在处理长序列数据时存在的梯度消失问题。它的基本思想是用 LSTM 对帧的 CNN 最后一层的激活在时间轴上进行整合。一方面,它可以对 CNN 特征进行更长时间的融合,不对处理的帧数加以上限,从而能对更长时长的视频进行表达;另一方面,不用考虑同一次进网络的帧的前后顺序。LSTM 的隐藏层的输出的具体计算过程如式(1)~式(6)所示:

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{c} = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (4)$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(c_t) \quad (6)$$

其中, w 表示连接两层的权重矩阵, b 表示偏置向量, c 表示记忆单元的状态, σ 和 \tanh 表示两种不同的神经元激活函数, i , f 和 o 分别表示输入门、遗忘门和输出门。然而 LSTM 只捕获了图片的上文信息,而对下文信息却一无所知,而提出的 BLSTM 可以同时考虑到数据信息的上下文信息。如图 5 所示,为 BLSTM 的网络架构图。

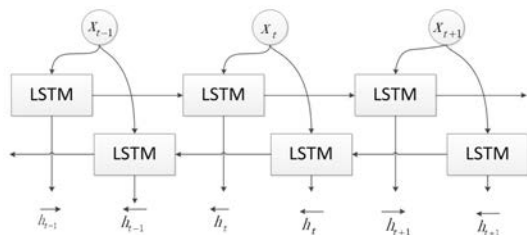


图5 BLSTM网络架构

本文提出通过 BLSTM 来学习不同人脸表情在时域上的关联特征。我们构建了多个 BLSTM 层,以及一个全连接层和一个 softmax 层。用已训练好的 CNN 模型来提取面部表情,在两层卷积降维之

后的输出作为提取的特征。随后依次输入到 BLSTM 网络中去,只有 t 时刻, BLSTM 才会输出它的特征到全连接层,最终送入到 softmax 层得到分类结果。

3 实验

软件环境: Windows 10 下的 TensorFlow 平台;

硬件环境: 处理器: Intel (R) Core (TM) i7-7700 CPU @ 3.60GHz; 内存: 32.0GB。

3.1 实验数据集

3.1.1 JAFFE 数据集

JAFFE (The Japanese Female Facial Expression Database) 数据库表情标定标准, 广泛用于面部表情识别验证中, 其中包含了 213 幅日本女性的脸相, 每幅图像都有原始的表情定义。表情库中共有 10 个人, 每个人有中性、高兴、悲伤、惊奇、愤怒、厌恶、恐惧 7 种表情。

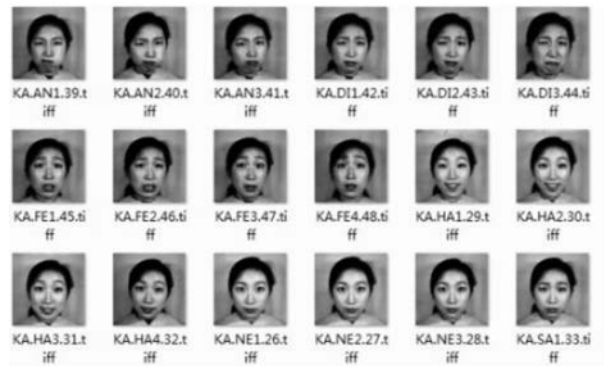


图6 JAFFE数据库中部面部表情示例

3.1.2 CK+数据集

这个数据库是在 Cohn-Kanade Dataset 的基础上扩展来的, 发布于 2010 年。包含表情的 label 和 Action Units 的 label。这个数据库包括 123 个 subjects, 593 个 image sequence, 每个 image sequence 的最后一张 Frame 都有 action units 的 label, 而在这 593 个 image sequence 中, 有 327 个 sequence 有 emotion 的 label。如图 7 所示为 CK+数据集的部分图片示例。



图7 CK+数据库中部面部表情示例

3.1.3 数据集的融合

由于每种数据集都有数据量、条件等多方面限

制,提出利用JAFfE数据集和CK+数据集融合的混合数据集。利用这两个数据集选取连续多帧图像并通过反转、裁剪等手段对数据集大小进行扩充,同时每次实验随机按比例选取数据集中部分图像作为测试数据,而其他的图像作为训练数据。实验重复5次,最后对实验结果取平均值得到总识别率。

对新的数据集进行整理和制作emotion的lables。并将它们按照比例无顺序拆分成训练样本和测试样本,具体分布如表2所示。

表2 不同表情的样本数

表情样本	训练样本数量	测试样本数量
中性	108	12
愤怒	612	60
厌恶	732	84
恐惧	372	48
高兴	852	96
悲伤	408	48
惊讶	1020	108

3.2 实验结果分析

将整理好的数据集作为实验数据输入至3D CNN网络中进行训练,获得特征数据之后输入至BLSTM网络中,并用softmax分类器进行分类得到分类结果。

在实验过程中,使用多种模型对三种数据集中图像进行表情识别。用到的模型主要有PCA、LBP、CNN、3D CNN和改进的CNN。如表3~5所示,为在三种数据集下,多种传统的模型的面部表情识别结果。

表3 利用JAFfE数据集识别准确率对比

模型	中性	愤怒	厌恶	恐惧	高兴	悲伤	惊讶
PCA	38.1	50.4	27.1	31.3	69.3	38.4	39.5
LBP	42.1	69.3	51.2	43.8	75.2	50.2	39.5
CNN	68.2	70.3	50.9	50.8	80.5	50.5	48.4
3D	72.5	80.5	66.4	75.7	92.4	78.4	65.6

表4 利用CK+数据集识别准确率对比

模型	中性	愤怒	厌恶	恐惧	高兴	悲伤	惊讶
PCA	38.1	50.4	27.1	31.3	69.3	38.4	39.5
LBP	42.1	69.3	51.2	43.8	75.2	50.2	39.5
CNN	68.2	70.3	50.9	50.8	80.5	50.5	48.4
3D	72.5	80.5	66.4	75.7	92.4	78.4	65.6

为了进一步证明本文采用的网络模型的有效性,我们在多种数据集测试数据集上依次重复测试5次求其平均值作为最终测试结果,使得最终的结果具有可信性。如表6~8所示。

表5 利用融合新的数据集识别准确率对比

模型	中性	愤怒	厌恶	恐惧	高兴	悲伤	惊讶
PCA	38.1	50.4	27.1	31.3	69.3	38.4	39.5
LBP	42.1	69.3	51.2	43.8	75.2	50.2	39.5
CNN	68.2	70.3	50.9	50.8	80.5	50.5	48.4
3D	72.5	80.5	66.4	75.7	92.4	78.4	65.6

表6 JAFfE下新模型表情识别准确率

	中性	愤怒	厌恶	恐惧	高兴	悲伤	惊讶
1	79.1	86.3	76.2	80.1	95.2	79.1	84.3
2	80.2	84.2	75.1	79.5	91.3	78.2	86.4
3	81.2	88.1	75.3	79.1	93.6	79.4	84.6
4	78.1	87.3	77.2	81.3	97.2	77.9	85.2
5	79.2	87.5	78.1	77.2	94.3	72.3	81.3

表7 CK+下新模型表情识别准确率

	中性	愤怒	厌恶	恐惧	高兴	悲伤	惊讶
1	72.3	87.2	79.1	71.7	92.8	71.4	81.1
2	71.5	89.3	78.4	72.3	93.1	70.5	79.6
3	75.4	87.5	80.3	69.8	93.2	70.8	80.3
4	72.4	88.4	77.5	71.6	91.8	73.6	80.6
5	73.8	86.9	79.7	70	93.5	69.5	81.5

表8 融合数据集下新模型表情识别准确率

	中性	愤怒	厌恶	恐惧	高兴	悲伤	惊讶
1	78.3	85.4	77.1	80.1	94.3	77.3	82.5
2	77.2	87.9	76.8	79.5	95.0	75.8	83.1
3	81.2	86.5	77.5	77.8	96.1	79.8	83.8
4	80.3	88.8	75.9	77.6	95.7	76.9	84.4
5	78.0	88.9	74.7	78.0	96.4	79.7	84.7

由表6~8可知,无论在何种数据集的情况下,3D CNN+BLSTM模型都表现出了良好的识别效果,这表明该模型具有较强的稳定性和识别性能。使用融合模型在多种数据集的面部表情识别率如图8所示。

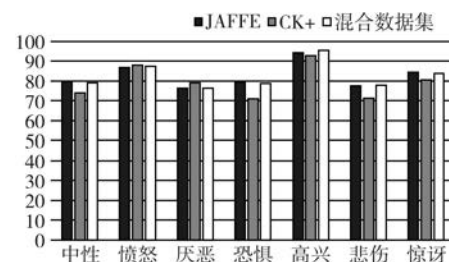


图8 新模型下各种数据集对比

如表9所示为新模型与传统模型识别率的对比。

由表9可知,与传统模型相比,新型融合模型在三种数据集下都表现出了良好的识别率。与PCA模型相比,总体识别率提高了50.3%,与LBP

模型相比,总体识别率提高了38%,与CNN模型相比,总体识别率提高了25.7%,与CNN模型相比,总体识别率提高了8.5%。

表9 新模型和传统模型识别率对比

数据集	PCA	LBP	CNN	3D CNN	3D CNN +BLSTM
JAFFE	40.1	49.7	62.5	75.8	83.3
CK+	42	53	59.9	75.9	81.7
混合	41.9	50.8	61.7	75	82.7

4 结语

本文提出采用融合卷积神经网络和递归神经网络的方法来对人的面部表情进行识别。该方法通过采用3D CNN对连续的多张图片中的特征进行提取,进而采用BLSTM递归神经网络得到这些特征在时域上的信息,进而对这些信息进行分类。实验数据表明,该方法的识别结果确实优于传统方法。但当数据量越大,网络层次较深的情况下,由于硬件设备的限制,模型的训练时间就会增长。在提高准确率的情况下,模型训练的时效性是进一步研究的重点。

参考文献

- [1] 万川. 基于动态序列图像的人脸表情识别系统理论与方法研究[D]. 长春:吉林大学, 2013.
WAN Chuan. Research on the theory and method of facial expression recognition system based on dynamic sequence image [D]. Changchun: Jilin University, 2013.
- [2] Nazima kausar, Jitendra Sharma. Apaten. Automatic Facial Expression Recognition: A Survey Based on Feature Extraction and Classification Techniques.[C]//International Conference on ICT in Business Industry & Government, 2016.
- [3] G. Chetty, M. Wagner, and R. Goecke. A Multilevel Fusion Approach for Audiovisual Emotion Recognition[J]. John Wiley & Sons, Inc., 2015:437-460.
- [4] Y. Kim, H. Lee, and E. Provost. Deep learning for robust feature generation in audiovisual emotion recognition[C]//In Acoustics, Speech and Signal Processing, 2013 IEEE International Conference on, 2013:3687-3691.
- [5] A. Konar, A. Halder, and A. Chakraborty. Introduction to Emotion Recognition[J]. John Wiley & Sons, Inc., 2015: 1-45.
- [6] Y. Tongand, Q. Ji. Exploiting Dynamic Dependencies Among Action Units for Spontaneous Facial Action Recognition[J]. John Wiley & Sons, Inc., 2015:47-67.
- [7] J. Schmidhuber. Deep learning in neural networks: An overview[J]. Neural Netw, 2015, 61:85-117.
- [8] Y. Lv, Z. Feng, and C. Xu, Facial expression recognition via deep learning[J]. IETE Tech. Rev., 2015, 32(5): 347-355.
- [9] Bishwas Mishra; Steven L. Fernandes. Facial expression recognition using feature based techniques and model based techniques [C]//A survey 2015 2nd International Conference on Electronics and Communication Systems, 2015.
- [10] Ganesh R. Naik, Dinesh K. Kumar. Limitations and Applications of ICA in Facial sEMG and Hand Gesture sEMG for Human Computer Interaction[C]//9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications, 2007.
- [11] Tanapol Pumlumchiak, Sirion Vittayakorn. Facial expression recognition using local Gabor filters and PCA plus LDA [C]//9th International Conference on Information Technology and Electrical Engineering, 2017.
- [12] Nazima Kausar, Jitendra Sharma. Facial expression recognition using LBP template of facial parts and multilayer neural network [C]//International Conference on I-SMAC, 2017.
- [13] K. Mase. Recognition of facial expression from optical flow [J]. IEICE Trans. Inf. Syst., 1991, E74-D(10): 3474-3483.
- [14] 徐姗姗. 卷积神经网络的研究与应用[D]. 南京:南京林业大学, 2013.
XU Shanshan. Research and application of convolutional neural network [D]. Nanjing: Nanjing Forestry University, 2013.
- [15] 李勇, 林小竹, 蒋梦莹. 基于跨连接LeNet-5网络的面部表情识别[J]. 自动化学报, 2018, 44(1):176-182.
LI Yong, LIN Xiaozhu, JIANG Mengying. Facial expression recognition based on cross-connected lenet-5 network [J]. Acta automata sinica, 2018, 44(1):176-182.
- [16] Sakshi Gupta, Ravindra K. Singh. Mathematical morphology based face segmentation and facial feature extraction for facial expression recognition [C]//2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management. IEEE, 2015.
- [17] Pukhrambam M, Das A, Saha A. Facial components extraction and expression recognition in static images [C]//International Conference on Green Computing and Internet of Things. IEEE, 2016.
- [18] Zarbakhsh P, Demirel H. Fuzzy SVM for 3D facial expression classification using sequential forward feature selection [C]//International Conference on Computational Intelligence and Communication Networks. IEEE, 2018.

- [19] Chen J, Bai Y. Classification of smile expression using hybrid PHOG and Gabor features[C]//International Conference on Computer Application and System Modeling. IEEE, 2010.
- [20] He J, Cai J F, Fang L Z, et al. A method of facial expression recognition based on LBP fusion of key expressions areas [C]//Control and Decision Conference. IEEE, 2015.
- [21] Luo Y, Zhang L, Chen Y, et al. Facial Expression Recognition Algorithm Based on Reverse Co-Salient Regions (RCSR) Features[C]//International Conference on Information Science and Control Engineering. IEEE Computer Society, 2017.
- [22] W. Liu, W. Zheng, and B. Lu. Multimodal emotion recognition using multimodal deep learning[J]. CoRR, abs/1602.08225, 2016:521-529.

(上接第2900页)

- [4] Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge [C]//ACM: Sigmod Conference, 2008. 1247-1250.
- [5] Lehmann J, Isele R, Jakob M, et al. DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia[J]. Semantic Web, 2015, 6(2): 167-195.
- [6] Suchanek F M, Kasneci G, Weikum G. Yago: a core of semantic knowledge[C]//Proceedings of the 16th international conference on World Wide Web, 2007: 697-706.
- [7] Bordes A, Weston J, Usunier N. Open question answering with weakly supervised embedding models [C]//Springer-Verlag New York, Inc: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2014: 165-180.
- [8] Ernst P, Meng C, Siu A, et al. Knowlife: a knowledge graph for health and life sciences[C]//IEEE: IEEE International Conference on Data Engineering, 2014: 1254-1257.
- [9] Ali M, Hoyt C T, Domingo-Fernández D, et al. BioKEEN: a library for learning and evaluating biological knowledge graph embeddings[J]. Bioinformatics, 2019, 35(18): 3538-3540.
- [10] Shi L, Li S, Yang X, et al. Semantic health knowledge graph: Semantic integration of heterogeneous medical knowledge and services [J]. BioMed research international, 2017: 1-12.
- [11] Heim P, Hellmann S, Lehmann J, et al. RelFinder: Revealing relationships in RDF knowledge bases[C]//International Conference on Semantic and Digital Media Technologies. Springer, Berlin, Heidelberg, 2009: 182-187.
- [12] Kerdjoudj F, Curé O. RDF knowledge graph visualization from a knowledge extraction system [J]. ArXiv preprint arXiv:1510.00244, 2015.
- [13] Domingo-Fernandez D, Baksi S, Schultz B, et al. COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology[J]. BioRxiv, 2020.
- [14] 中国医学科学院, 北京协和医院. 协和新型冠状病毒肺炎防护手册[M]. 北京: 中国协和医科大学出版社, 2020: 24-88.
- Peking Union Medical College Hospital, Chinese Academy of Medical Sciences. Concord New Coronavirus Pneumonia Protection Manual [M]. Beijing: Peking Union Medical University Press, 2020: 24-88.
- [15] 张文宏. 张文宏教授支招防控新型冠状病毒[M]. 上海: 上海科学技术出版社, 2020: 4-50.
- ZHANG Wenhong. Professor Zhang Wenhong advises on the prevention and control of new coronaviruses [M]. Shanghai: Shanghai Science and Technology Press, 2020: 4-50.
- [16] 中央赴湖北省指导组防控组. 不同人群、不同场所和不同交通工具健康防护指导手册[EB/OL]. 2020-02-15, http://www.hubei.gov.cn/zhuanti/2020/gzxxgzbd/fkkp/202002/t20200215_2028524.shtml.
- The Prevention and Control Team of the Central Guidance Group to Hubei Province. Guidance Manual for Health Protection of Different Groups, Different Places and Different Vehicles [EB/OL]. 2020-02-15, http://www.hubei.gov.cn/zhuanti/2020/gzxxgzbd/fkkp/202002/t20200215_2028524.shtml.