

文章编号: 2095-2163(2022)02-0013-05

中图分类号: TP391.41

文献标志码: A

基于联合损失函数的多视角步态识别方法研究

赵佳鑫, 张雅丽, 施新凯, 李御瑾

(中国人民公安大学 信息网络安全学院, 北京 100038)

摘要: 视频侦查技术逐步成为侦查办案的重要手段, 步态识别技术在视频侦查工作中有良好的应用前景; 现阶段跨视角的步态识别技术在公开数据集 CASIA-B 的正常行走条件下, 识别正确率已经达到了 95%, 但是在穿外套和携带包裹等有遮挡行走条件下的识别正确率不高。本文通过采用 Triplet 与 Softmax 两个损失函数相结合的方法, 联合优化步态模型, 可以使最终的训练更加有效。研究结果表明: 在不影响正常行走条件识别正确率的情况下, 穿外套和携带包裹行走条件的步态识别正确率均有一定提升, 分别提升了 4.7% 和 3%, 可为视频目标识别与追踪提供辅助技术支持。

关键词: 跨视角; 步态识别; Triplet; Softmax

Research on multi-view gait recognition method based on joint loss functions

ZHAO Jiabin, ZHANG Yali, SHI Xinkai, LI Yujin

(Information Technology & Network Security Institute, People's Public Security University of China, Beijing 100038, China)

[Abstract] Video investigation technology has gradually become an important means of investigating and handling cases, and gait recognition technology has a good application prospect in video investigation work; Under the normal walking conditions of the CASIA-B public dataset, the correct of cross-view gait recognition has reached 95% at this stage, but the recognition correct is not satisfactory under occluded walking conditions such as wearing a coat and carrying a package. By using the method of combining the two loss functions of Triplet and Softmax, and jointly optimizing the gait model, the final training can be more effective. The results of the study show that, without affecting the recognition correct of normal walking conditions, the correct of gait recognition in the conditions of wearing a coat and walking with a package has been improved to a certain extent, increasing by 4.7% and 3% respectively. The research results can provide auxiliary technical support for video target recognition and tracking.

[Key words] cross-view; gait recognition; Triplet; Softmax

0 引言

视频侦查技术具有可以突破时空限制、证据固定等优势, 逐步成为公安办案的新手段^[1]。在视频侦查工作中, 目标识别与追踪工作一直是案件侦破的重要环节, 其效率与准确性至关重要。犯罪分子因其反侦查意识越来越强, 在作案过程中往往是全副伪装, 使得很多技术手段受限^[2]。现有的人脸识别技术主要应用在人像卡口或装有高清监控设备的场景下, 犯罪嫌疑人实施犯罪的过程中会故意利用摄像头盲区、伪装人脸图像信息等致使身份无法识别^[3]。而步态识别可以利用现有摄像头进行远距离识别, 嫌疑人难以隐藏与伪装, 在视频侦查中具有明显的优势, 成为近年来身份识别领域研究的热门。

步态识别是一种分析行人走路的姿态, 利用算法将其从步态数据库中识别出来的方法, 已经在行

人无更换外衣、无携带、摄像方向无改变的情况下, 具有较高的识别正确率, 然而面对复杂的真实场景, 该技术仍有许多不容易解决的问题, 如: 行走路面不平 and 摄像头角度变化会产生视角问题, 有无穿外套和是否携带包裹产生遮挡问题, 这些因素导致其在复杂的实际行走过程中识别正确率仍然较低^[4]。

本文主要研究基于深度学习的步态识别技术, 在 GaitSet 模型中利用 Triplet 和 Softmax 联合损失函数来提升行人在穿大衣和携带包裹这两种情况下的步态识别正确率。

1 研究方法

1.1 GaitSet 网络

识别步态有两种主要方法, 第一种方法将步态视为图像, 典型的代表是提取步态能量图^[5]。步态能量图是步态识别任务中最常用的特征图, 其主要

基金项目: 中国人民公安大学 2021 年度基本科研业务费重点项目(2021JKF2051)。

作者简介: 赵佳鑫(1996-), 男, 硕士研究生, 主要研究方向: 智能视频分析; 张雅丽(1977-), 女, 硕士, 副教授, 主要研究方向: 智能视频分析。

通讯作者: 张雅丽 Email: zhangyl_mail@163.com

收稿日期: 2021-11-18

工作是把一组剪影图对齐压缩成一张图片,利用步态模板对行人进行分类,通过计算在长时间范围内的平均步态剪影,可以减少在预处理阶段中由于处理不完全而残留的噪点。步态能量图是当前最常用的步态识别方法之一,能够实现计算量与识别准确率相对均衡,一直以来都被当作步态识别的代名词。但这种方法没有考虑到步态中的时序信息和空域信息,其步态识别正确率并不高。

第二种方法将步态视为视频序列来处理,直接从最初的剪影序列中提取步态特征,常使用 LSTM 方法、3D-CNN 方法或者双流法(two stream)。视频信号和语音信号都含有时序信息,而 LSTM 模型通常用于处理这些带有时间顺序的信息,首先利用卷积神经网络提取每帧视频的步态特征,然后利用 LSTM 对时序关系建模;3D-CNN 模型可以处理视频比图像多出的时间维度,常被用于视频分类工作,然而由于其计算任务重以及计算周期较长,往往难以在实际中得到真正的应用。

双流法是包含 RGB 图像和光流两个通道的视频行为识别方法。其中,光流通道被用来建模时序信息,RGB 图像通道被用来建模空域信息,该方法将两个通道进行信息融合,并联合训练,可以很好的

建模步态序列中的时序和空域信息,但其往往会受限于一些外部因素。

为了兼顾运算量与识别正确率,Chao 等人提出了一套新的解决方案:把步态特征看作一组由视频帧组成的图像序列。行走是一种周期性的运动,所以步态可以选用其中一个周期进行表示。在一个周期的步态序列中,可以观察到不同位置的剪影,都具有可以区别于其他剪影的外形,如图 1 所示。



图 1 一个周期的步态剪影

Fig. 1 A cycle of gait silhouette

仅需通过观察步态剪影的大致形状,便能一一分辨,即使面对已经被全部打乱顺序的步态剪影图,也能再次组成其原本正确的顺序。因此,假定步态剪影的外形本身就包含着顺序,进而剪影图的先后信息将不再是必要的输入,能够直接把步态视为一组图像来提取时序信息。Chao 等人在此理论基础上,提出了一种 GaitSet 深度学习网络模型,添加联合损失函数后网络框架如图 2 所示^[6]。

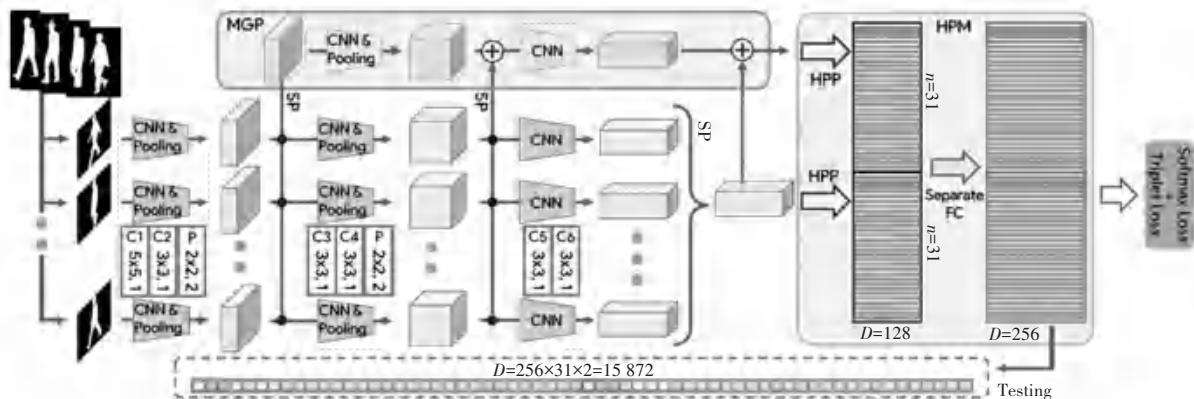


图 2 添加联合损失函数 GaitSet 框架图

Fig. 2 GaitSet frame diagram for adding joint loss function

本文采用 GaitSet 模型结构。GaitSet 模型十分灵活,输入仅会对轮廓图的大小加以限制,即输入的序列可以包含有不同角度、具有随意组合的行走状态、数量不受限的、可间断的剪影。该模型不计算一对步态序列间的相似度,而是直接学习步态的表示。因此,只需要对每个样本的表示计算一次,就能够通过测量不同样本的表示之间的欧式距离来进行分类识别。GaitSet 极大地提高了在 CASIA-B 和 OUMVLP 两大公开数据集的识别性能,显示了其在

不同视角和多行走条件下的强大鲁棒性以及大型数据集的高泛化能力。

1.2 联合损失函数

联合损失函数将多种损失函数结合,起到联合优化的效果。本文采用 Softmax loss 和 Triplet loss 两个损失函数联合训练的方法,增大步态示例类间距离,缩小步态示例类内距离,使得模型在多个损失函数的优化下,提高表征能力。

Triplet 损失函数通常用于训练差异性较小的示

例,其具体公式(1):

$$L_i = \sum_i^N [d(a,p)^2 - d(a,n)^2 + \text{margin}]_+ \quad (1)$$

其中, L_i 表示损失; a 表示锚(anchor) 示例; p 表示正(positive) 示例; n 表示负(Negative) 示例; N 表示一组 Batch 中的所有示例个数。

Softmax 损失函数最常用于处理图像多分类任务,其具体公式(2):

$$L_s = -\frac{1}{N} \sum_i^N \log \frac{\exp(\mathbf{W}_{y_i}^T \mathbf{X}_i + b_{y_i})}{\sum_j^C \exp(\mathbf{W}_j^T \mathbf{X}_i + b_j)} \quad (2)$$

其中, L_s 表示损失; \mathbf{X}_i 表示第 i 个步态图像的特征向量; y_i 是 \mathbf{X}_i 的真实标签; b 为偏置; $\mathbf{W}_{y_i}^T$ 与 \mathbf{W}_j^T 分别表示将 \mathbf{X}_i 判别为 y_i 类和 j 类的权向量; C 表示总类别数; $\mathbf{W}_{y_i}^T \mathbf{X}_i + b_{y_i}$ 表示步态图像在类别 y_i 上的得分; N 表示总体训练样本数。

最终的联合损失函数公式(3):

$$\text{Loss} = \alpha L_t + \beta L_s \quad (3)$$

其中, α 和 β 分别代表 Triplet 和 Softmax 两个损失函数的权重; 当 $\alpha = 0$ 且 $\beta \neq 0$ 时为 Softmax 损失函数; 当 $\beta = 0$ 且 $\alpha \neq 0$ 时为 Triplet 损失函数; 当 $\alpha \neq 0$ 且 $\beta \neq 0$ 时为联合损失函数。

2 实验

2.1 CASIA-B 数据集

CASIA-B 是 2005 年由中国科学院自动化研究所提出的免费公开人体步态数据集,该数据集是一个具有大规模、多角度、多行走条件等优势的数据集,在步态识别领域具有很高的地位^[7]。数据集包含 124 个不同身份的行人(ID 编号为 1~124),每个行人分别从 11 个视角($0^\circ \sim 180^\circ$)对普通行走(NM)、穿大衣(CL)和携带包裹(BG)这 3 种行走条件进行录制, NM 行走条件录制 6 次、CL 和 BG 两种行走条件各录制 2 次,视频中某个 ID 的行人在普通行走条件下,不同视角的步态图像例子如图 3 所示。



图 3 NM 类型步态图像示例

Fig. 3 Example of NM type gait image

2.2 数据预处理

每个数据集具有 13 640 个步态视频,同时拥有 1 364 个背景视频,总计 15 004 个,视频帧的长度也在几个到几百个不等,共占用 16 GB 的内存。而在实验时,在对数据集的信息进行整理分析的过程中,发现 026 和 027 两个行人 ID 的步态视频因为数据本身存在瑕疵而无法正常使用,同时剪影图中行人轮廓在 109 号视频中也基本没有出现,所以能用到的视频总共有 14 614 个,步态视频约占 91%,其余各类视频占 9%。数据经过预处理之后,产生人体步态图像约 1 116 300 张,占用内存 3.2 GB 左右。预处理后,行人 ID 为 010,在携带包裹条件下,01 次录制,视角为 90° 的剪影图像如图 4 所示;预处理后,行人 ID 为 010,在穿大衣条件下,01 次录制,视角为 90° 的剪影图像如图 5 所示。



图 4 CASIA-B 数据集预处理后携带包裹剪影图像

Fig. 4 BG silhouette image after preprocessing of CASIA-B data set



图 5 CASIA-B 数据集预处理后穿大衣剪影图像

Fig. 5 CL silhouette image after preprocessing of CASIA-B data set

2.3 模型训练

在 GaitSet 网络模型的基础上,将 CASIA-B 数据集采用大样本训练的划分形式。按照惯例,所划分的训练集部分要完全输入到 GaitSet 网络进行训练。但在验证的过程中,由于所有行人中的部分步态图像需要作为对照组,所以另一部分未被作为对照而是用作识别的步态图像被称为探针。因为要均衡 NM、BG、CL 三者之间的量,防止其中某一行走条件的信息过多或者过少,所以实验过程中,将所有行人 ID 中的 nm-05、06、cl-01、bg-01、cl-02、bg-02 剪影图像用作识别,剩余的 nm-01、02、03、04 剪影图像组成步态图库。

本次实验使用 8 个 NVIDIA Tesla K80 12 G 显卡,全连接隐层数量为 256,学习率为 0.000 1,视频提取帧数为 30,批尺寸为 8×16 ,三元组损失 Margin 为 0.2,训练轮数 80 000。

在训练过程中,把 100 轮分为一组,共训练 800 组,每组的平均损失函数如图 6 所示。可以观察到,与单独使用 Triplet 损失函数单调递增趋势不同的是,联合损失函数整体呈现单调递减趋势。由于训

练的所有 Batch 数据都是由随机目标 ID 的随机类型、随机角度和随机帧序列等组成,所以每个相邻 Step 之间都有一定程度的震荡,因此损失值最终能收敛到 margin 值附近即合理。

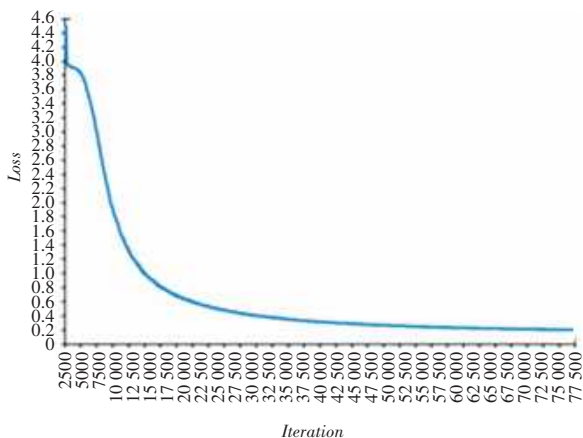


图 6 损失函数曲线

Fig. 6 Loss function curve

3 实验结果分析

3.1 联合损失函数权重取值

在联合损失函数权重的设置中,公式(3)中 α 和 β 的不同取值,会影响模型的最终识别效果,为了达到在穿大衣和携带包裹两种情况下较高的识别正确率,需选取最佳的 α 值和 β 值,本文对 α 和 β 分别采用 1:0、0.9:0.1、0.8:0.2、0.7:0.3、0.6:0.4、0.5:0.5、0.4:0.6、0.3:0.7、0.2:0.8、0.1:0.9、0:1,共 11 组不同的权重分配方法进行实验,对比结果如图 7 所示。

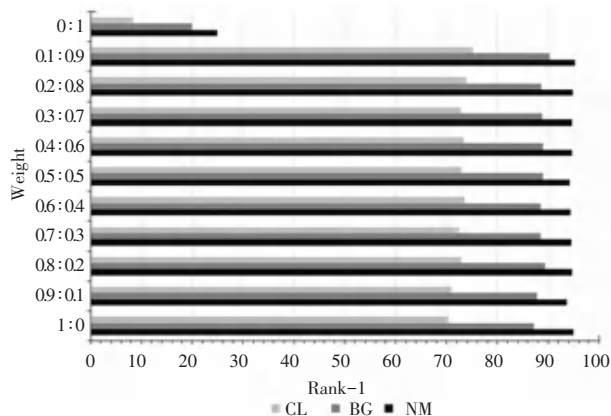


图 7 两类损失函数在不同权重下的测试结果对比

Fig. 7 Comparison of the test results of the two types of loss functions under different weights

当 $\alpha = 0$ 且 $\beta = 1$ 时,表示单独使用 Softmax 损失函数训练 GaitSet,精确度曲线如图 8 所示。

可以看出,虽然在单独使用 Softmax 损失函数训练 GaitSet 网络时,最终分类精确度很高,但是在

图 7 的测试结果中,采用 $\alpha = 0, \beta = 1$ 权重训练得到的模型,在 3 种行走状态下的识别正确率都很低,可见 Triplet 损失函数是必不可少的;而在 Triplet 损失函数和 Softmax 损失函数联合训练的测试结果中,穿大衣和携带包裹两种行走条件在所有权重下识别正确率均有不同幅度的提升,其中 0.1:0.9 权重的识别正确率提升最大;而在 0.2:0.8、0.3:0.7、0.4:0.6、0.5:0.5、0.6:0.4、0.7:0.3、0.8:0.2、0.9:0.1 这几种权重的测试结果中,普通行走条件下的识别正确率相较于 GaitSet 原网络均有小幅下降,唯有 0.1:0.9 权重的识别正确率有小幅上升,所以选定 $\alpha = 0.1, \beta = 0.9$,最终损失函数如公式(4)所示。

$$Loss = 0.1L_t + 0.9L_s \quad (4)$$

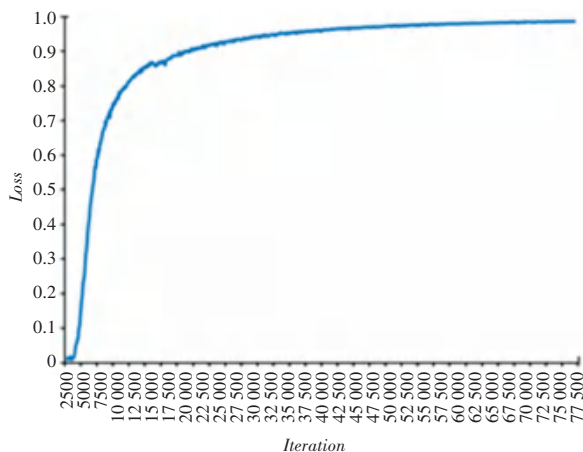


图 8 仅使用 Softmax loss 训练 GaitSet 的精确度曲线

Fig. 8 Only use the Softmax loss to train the accuracy curve of GaitSet

3.2 实验结果分析

在 LT 的训练集划分形式下,将本文的实验结果根据行走条件同 GaitSet 和 CNN-LB 进行对比,记录在 11 个不同视角下的识别正确率,并分别计算每种行走条件下的所有视角识别正确率的平均值,结果如表 1 所示。

表 1 中 3 种算法在正常行走条件下的识别情况差别不大,识别正确率都高于 91%,尤其是 GaitSet 算法和本文的方法均已高达 95%;在携带包裹的行走条件下,可以推测三种算法都因行人身体被部分遮挡导致识别正确率受到影响,但相对来说,受影响最小的是本文利用联合损失函数训练 GaitSet 网络的方法,步态识别正确率突破 90%;三种算法对于行人因穿大衣而导致身体被大面积遮挡的情况,识别效果都不理想,而本文的方法达到了 75.1% 的识别正确率,具有相对更好的识别结果。

表 1 多视角下的步态识别实验结果对比

Tab. 1 Comparison of experimental results of gait recognition under multiple viewing angles

Gallery NM#1-4			0°-180°											mean
probe			0	18	36	54	72	90	108	126	144	162	180	
LT(74)	NM#5-6	CNN-LB	87.7	92.0	95.3	94.2	89.9	86.5	90.2	95.0	96.5	92.9	82.9	91.2
		GaitSet	90.8	97.9	99.4	96.9	93.6	91.7	95.0	97.8	98.9	96.8	85.8	95.0
		ours	89.7	97.9	99.1	97.4	93.4	92.3	95.2	98.1	98.2	98.1	88.0	95.2
BG#1-2	NM#5-6	CNN-LB	64.2	80.6	82.7	76.9	64.8	63.1	68.0	76.9	82.2	75.4	61.3	72.4
		GaitSet	83.8	91.2	91.8	88.8	83.3	81.0	84.1	90.0	92.2	94.4	79.0	87.2
		ours	85.0	94.5	93.3	91.8	88.3	84.4	89.7	91.9	94.9	94.5	84.4	90.2
CL#1-2	NM#5-6	CNN-LB	37.7	57.2	66.6	61.1	55.2	54.6	55.2	59.1	58.9	48.8	39.4	54.0
		GaitSet	61.4	75.4	80.7	77.3	72.1	70.1	71.5	73.5	73.5	68.4	50.0	70.4
		ours	67.5	80.4	82.1	77.7	75.5	70.3	74.8	77.2	80.0	77.3	63.3	75.1

4 结束语

在多视角的监控视频下,对携带包裹和穿外套的步态识别效果普遍不佳。针对上述不足,本文提出了一种 Triplet 损失与 Softmax 损失联合监督的步态识别方法,增强模型的泛化能力,在几乎不增加模型运算量的同时,提高了携带包裹和穿大衣这两种行走状态下的识别正确率。由于实际场景的复杂性,步态识别技术目前还难以直接应用于视频侦查实战。但是,当犯罪嫌疑人进行面部遮挡、伪装以及光照条件较差时,人脸识别技术将会受限,步态识别技术可以作为人脸识别的重要补充间接应用于公安实践中。

参考文献

[1] 张玮佳. 人脸识别在侦查中的应用与困境研究[J]. 轻工科技,

2021, 37(4): 114-115.

- [2] 朱应钊, 李嫒. 步态识别现状及发展趋势[J]. 电信科学, 2020, 36(8): 130-138.
- [3] 周悦媛. 基于步态特征的身份识别算法研究[D]. 哈尔滨: 哈尔滨工程大学, 2020.
- [4] 周石冰. 基于轮廓与姿态特征融合的步态识别研究[D]. 哈尔滨: 哈尔滨工程大学, 2020.
- [5] JU H, BHANU B. Individual recognition using gait energy image [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 28(2): 316-322.
- [6] CHAO H, HE Y, ZHANG J, et al. GaitSet: Regarding gait as a set for cross-view gait recognition [C]// Proc of the AAAI Conference on Artificial Intelligence, 2019, 33: 8126-8133.
- [7] YU S, TAN D, TAN T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition [C]// Proc of the 18th International Conference on Pattern Recognition (ICPR), 2006: 441-444.

(上接第 12 页)

- [7] 王晗, 孙雨晨. 基于交易三阶段的电子商务卖家信用评价模型研究[J]. 征信, 2020, 38(3): 31-38.
- [8] 刘潇雅, 王应明. 基于支持向量机集成的个人信用评估研究[J]. 计算机工程与应用, 2020, 56(14): 272-278.
- [9] 江义火, 袁晓建. 电子商务交易中信用度准确评价仿真研究[J]. 计算机仿真, 2018, 35(7): 430-433.
- [10] 吴澎, 周礼刚, 陈华友. 基于语言共识模型的电子商务信用风险评价方法[J]. 控制与决策, 2021, 36(6): 1465-1471.
- [11] 孙向宇, 晁涛, 王松艳, 等. 基于灰色层次属性模型的评估指标综合评价方法[J]. 系统仿真学报, 2018, 30(2): 699-706.
- [12] 杜志平, 张盟. 跨境电商物流企业绩效评价[J]. 统计与决策, 2020, 36(21): 181-184.
- [13] 杨韵. C2C 交易中的动态信用评价模型[J]. 情报科学, 2010, 28(4): 563-566.
- [14] 李怀栋. B2C 农产品电子商务信用评价指标体系的构建[J]. 征信, 2019, 37(2): 45-49.

- [15] 何佳晓, 王胜. 电商平台农产品经营主体信用评价指标体系构建研究[J]. 征信, 2020, 38(8): 53-59.
- [16] 贾应丽. 中国旅游企业商务信用评估体系研究[D]. 中央财经大学, 2018.
- [17] DIAKOULAKI D, MAVROTAS G, PAPAYANNAKIS L. Determining Objective Weights in Multiple Criteria Problems: The Critic Method [J]. Computers & Operations Research (S0305-0548), 1995, 22(7): 763-770.
- [18] 郭亚军. 综合评价理论与方法[M]. 科学出版社, 2002.
- [19] OPRICOVIC S, TZENG G H. Compromise Solution by MCDM Methods: A Comparative Analysis of VIKOR and TOPSIS [J]. European Journal of Operational Research (S0377-2217), 2004, 156(2): 445-455.
- [20] 刘丹. 基于熵权的 TOPSIS 生鲜电商企业供应商选择研究[J]. 管理科学与工程, 2019, 8(1): 1-6.