

密级:_____



中国科学院大学
University of Chinese Academy of Sciences

硕士学位论文

基于主题模型的视频集群目标轨迹分类

作者姓名: _____ 崔妍婷 _____

指导教师:叶齐祥 副教授 中国科学院大学电子电气与通信工程学院

学位类别: _____ 工程硕士 _____

学科专业: _____ 计算机技术 _____

研究所 : _____ 中国科学院大学工程管理与信息技术学院 _____

二〇一五年五月

Trajectory Classification for Crowded Scenes via
Topic Models

By
Yanting Cui

A Thesis Submitted to
The University of Chinese Academy of Sciences
In partial fulfillment of the requirement
For the degree of
Master of Computer Technology

College of Engineering and Information Technology
University of Chinese Academy of Sciences

May, 2015

中国科学院大学直属院系 研究生学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

日 期：

中国科学院大学直属院系 学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密的学位论文在解密后适用本声明。

作者签名：

日 期：

导师签名：

日 期：

摘 要

近年来，随着恐怖袭击事件频频发生，恐怖主义势力活动猖獗，公共场所的安保问题受到了全世界范围内的关注。公共场所中运动目标行为的监控、分析、预测以及对异常行为和突发事件的报警，对于国家社会的公共安全有着不言而喻的重要意义。然而现有的对公共场所的监控工作，主要是通过事后相关工作人员对监控视频进行观看和分析来寻找事件引发的可疑人员。随着经济发展和社会的进步，监控数据的爆炸式增长决定了无法通过事后人工分析的方式对公共场所进行有效的监控，同时也无法达到实时监控的效果。由此可见，对集群场景的运动目标行为分析的需求十分迫切。

本文完成的主要工作包括：

(1) 充分考虑集群视频场景中目标轨迹不完整、混杂的特点，提出 Global MRF (Global Markov Random Field, 全局马尔科夫随机场) 对原始轨迹进行预处理；

(2) 充分考虑轨迹表达的准确程度和框架的可扩展性，提出采用 BCTM (Belief based Correlated Topic Model, 基于置信关联主题模型) 挖掘轨迹的中层特征并在此基础上完成轨迹的重描述；

(3) 借鉴贝叶斯模型判别扩展的思想，利用多类 SVM (Multiclass Support Vector Machine, 多类支持向量机) 分类器，设计合理的分类流程，有效提高轨迹分类准确性；

(4) 借鉴贝叶斯模型生成扩展的思想，提出 Multi-CTM (Multi-Correlated Topic Model, 多关联主题模型) 用于轨迹中层特征的学习和分类。充分考虑到轨迹分类研究工作中人工标定带来的代价，设计并实现基于 Multi-CTM 的弱监督轨迹分类方法。

关键词： 集群场景，视频监控，轨迹分类，主题模型

Abstract

With the increase of terrorist attacks, the security issues in public places have drawn more attentions worldwide. Effective behavior monitoring, analysis, and prediction for moving objects in public places are of great significance. However, the existing work are mainly related to afterwards event retrieval. With the population of big data of surveillance video, the demand of real time behavior analysis of moving objects in crowded scenes is very urgent.

The primary works in this thesis are as follows:

1. By taking a full consideration of the challenge of fragmentary and chaotic trajectories, Global-MRF is proposed to preprocess trajectories.

2. By taking a full consideration of the accuracy of the trajectory representation and expansibility of framework, BCTM is used to learn middle level features of the trajectories.

3. Based on the idea of discriminant extension, the multi-SVM is used to design reasonable classified procedure, which can improve the precision of the classification.

4. According to the idea of generative extension, Multi-CTM is proposed to learn middle level features of the trajectories and trajectory classification. In order to reduce the hand-craft calibration, a method of Multi-CTM based weakly supervised trajectory classification is proposed.

Key Words: Crowded scene, Video surveillance, Trajectory classification, Topic model

目录

摘 要.....	I
Abstract.....	III
图目录.....	VII
表目录.....	IX
第一章 绪论	1
1.1 研究背景与意义.....	1
1.2 研究进展及存在的问题.....	4
1.2.1 非集群场景中运动目标轨迹分析.....	5
1.2.2 基于视觉的集群目标轨迹分析.....	9
1.3 本文主要研究内容.....	13
第二章 相关技术研究	15
2.1 视觉跟踪算法.....	15
2.2 MRF.....	16
2.3 BOW.....	18
2.3.1 文本 BOW 字典构建.....	18
2.3.2 图像 BOW 字典构建.....	19
2.3.3 轨迹表达中 BOW 字典构建.....	20
2.4 主题模型.....	21
2.4.1 LDA.....	22
2.4.2 CTM.....	27
2.5 本章小结.....	28
第三章 基于 BCTM 与多类 SVM 的集群目标轨迹分类	29
3.1 研究工作框架.....	29
3.2 场景结构.....	30
3.3 运动目标轨迹提取.....	32
3.4 轨迹预处理.....	33
3.4.1 Global MRF.....	33
3.4.2 实验结果与分析.....	35
3.5 中层特征学习.....	36
3.5.1 BCTM.....	38
3.5.2 模型公式推导.....	39
3.5.3 轨迹重描述.....	40
3.6 轨迹分类.....	41
3.6.1 多类支持向量机.....	41
3.6.2 分类流程的设计与实现.....	42
3.7 实验结果与分析.....	43
3.8 本章小结.....	47

第四章 基于 Multi-CTM 集群目标轨迹分类.....	49
4.1 主题模型与分类	49
4.2 classLDA	50
4.3 Multi-CTM	51
4.4 研究工作框架	52
4.4 实验结果与分析	54
4.5 本章小结	58
第五章 结论与展望	59
参考文献.....	61
附录：主要符号对照表.....	65
致 谢	67
个人简介、在学期间发表的论文与研究成果.....	69

图目录

图 1-1	传统视频监控系统.....	2
图 1-2	集群场景高俯角监控设备获取的图像.....	2
图 1-3	运动目标轨迹图.....	3
图 1-4	轨迹完整性统计图.....	3
图 1-5	非集群场景中轨迹分析流程图.....	5
图 1-6	监控场景中 POI 区域示意图.....	10
图 1-7	有结构集群场景和无结构集群场景.....	10
图 1-8	不同集群的自组织行为.....	12
图 1-9	本文整体框架图.....	13
图 2-1	片段轨迹时空一致性示意图.....	16
图 2-2	一副图像 SIFT 特征表达示意图.....	19
图 2-3	轨迹量化示意图.....	20
图 2-4	LDA 原理示意图.....	23
图 2-5	LDA 图模型 (a) 和近似图模型 (b).....	25
图 2-6	CTM 的图模型.....	28
图 3-1	本章研究工作框架图.....	29
图 3-2	场景结构图.....	31
图 3-3	场景进出口区域划分示意图.....	31
图 3-4	轨迹类别示意图.....	32
图 3-5	MRF 中轨迹绕圈图.....	34
图 3-6	Global MRF 轨迹预处理示意图.....	34
图 3-7	Global MRF 和 MRF 实验结果.....	36
图 3-8	片断轨迹被连接到不同的轨迹树中.....	37
图 3-9	BCTM 的图模型 (a) 和近似模型 (b).....	38
图 3-10	多类 SVM 与“One against All”对比图.....	42
图 3-11	纽约火车站数据集中层特征可视化实验结果.....	44
图 3-12	交通数据集中层特征可视化实验结果.....	45
图 3-13	分类准确率曲线图.....	46
图 4-1	classLDA 的图模型.....	50
图 4-2	不同类别的场景包含相同元素“天空”.....	51
图 4-3	multi-CTM 图模型.....	52
图 4-4	研究工作框架图.....	53
图 4-5	纽约火车站数据集中层特征可视化结果.....	57
图 4-6	交通场景数据集中层特征可视化结果.....	57

表目录

表 1-1	语料库中单词与索引对应表.....	18
表 4-1	交通场景数据集运动目标轨迹分类准确率.....	54
表 4-2	纽约火车站数据集运动目标轨迹分类准确率	55

第一章 绪论

本章包含三部分内容，首先介绍选题的背景和意义，然后介绍监控场景轨迹分类中相关工作和技术，最后概括本文的研究内容和论文的组织安排。

1.1 研究背景与意义

近年来，暴力恐怖袭击事件频频发生，恐怖主义势力不断猖獗，公共安全成为了全世界的焦点，各个国家对公共安全技术的研究投入了大量的人力和物力。有消息指出^[1]，2013 年全球发生 9707 起恐怖袭击，造成死亡人数多达 17800，分别较 2012 年增长 44% 和 61%。值得注意的是，暴力恐怖袭击往往发生在人员密集、社会关注程度高的大型群众性公共场所，如车站、机场、购物中心、中心广场等^[1]。因此，公共安全技术的研究已经成为了迫在眉睫的课题。

视频监控技术被广泛的应用到当今世界的各个角落，不论是室内还是室外，视频监控设备都成为了必不可少的安保设施。传统的视频监控技术可以称为人力视频监控，主要是依靠安保工作人员对监控设备获取的视频或者图像进行监看（见图 1-1）。由于视频监控设备数量的不断快速增长及其产生的海量监控视频数据，传统的视频监控技术需要付出的人力成本和时间成本越来越多，甚至可能超出了人力的处理范围。文献[2]指出：“一个接受过严格培训的视频监看人员最多能够有效地同时监看十六路监控视频，对于许多庞大的视频监控系统，当其监控设备数量很多时，监看视频的工作人员数量的增加并不一定能够提高监控的效率。同时，视频监看是一项非常疲劳的工作，长时间工作容易使监控者产生视觉疲劳，注意力下降，从而导致监控的效率和准确率大大降低，出现错报和漏报的现象”。面对海量的监控视频数据，我们需要的监控系统能够做到在几乎没有人力干预的情况下，每天连续 24 小时的实时智能监视，并能够对监控设备捕获的海量视频和图像数据自动地进行分析；当异常发生时，监控系统能够向安保人员发出及时、准确的警报信号，从而避免暴力恐怖事件的发生^[3]。因此，智能监控是一个不可阻挡的趋势。



图 1-1 传统视频监控系统

智能监控技术是一个跨学科的综合课题，其涉及图像处理、图像分析、计算机视觉、模式识别、机器学习等众多研究领域，同时也是一个极具挑战性的课题。运动目标轨迹分类作为实现智能监控技术的重要手段之一，主要是对监控场景中的运动目标进行定位、识别和跟踪，获取其运动轨迹，并在此基础上通过对轨迹数据进行建模完成对监控场景中运动目标的行为分类。人员密集、社会关注度高的公共场所，可以称之为集群场景。集群场景具有三个特征：

1) 集群场景中的监控设备一般呈现高俯角放置：这样的放置方式，只能远距离高倾斜角度对场景中运动目标进行拍照或者录像，无法获捕获到运动个体的静态特征（见图 1-2^[5]），如五官、面目表情、衣着等；只能捕获到运动个体的动态特征，如运动轨迹。因此，运动目标轨迹是集群场景运动目标分析可利用的有限数据之一。



图 1-2 集群场景高俯角监控设备获取的图像

2) 集群场景中人流密度大：运动目标之间遮挡严重，传统的跟踪识别算法捕获到的轨迹呈现高度片段性。图 1-3 是用 KLT (Kanade-Lucas-Tomasi, 关键点跟踪器)^[4]获取两个公开集群场景监控视频中运动目标轨迹图，其中 (a) 纽约火车站场景数据集^[5]，(b) 是交通场景数据集^[6]，从图中明显的看出，两个场景中运动目标轨迹呈现高度片断性和混乱性；为了对集群场景中运动目标轨迹的片断性有一个量的衡量，图 1-4 是对图 1-3 中目标轨迹完整性的统计图，“complete”，“only source”，“only sink”和“no source&sink”分别对应“完整轨迹”，“只有进口的轨迹”，“只有出口的轨迹”和“没有进口也没有出口的轨迹”，图 1-4 (a) 所示纽约火车站场景中完整目标轨迹数量占有所有目标轨迹数量 2.5%，图 1-4 (b) 所示交通监控场景中完整轨迹数量占不到 8%；

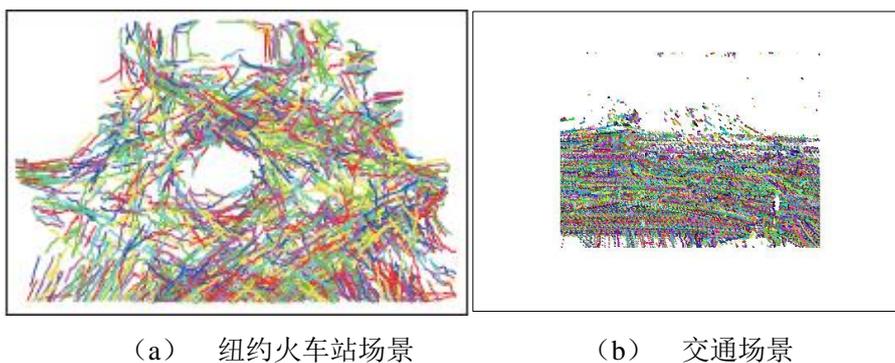


图 1-3 运动目标轨迹图

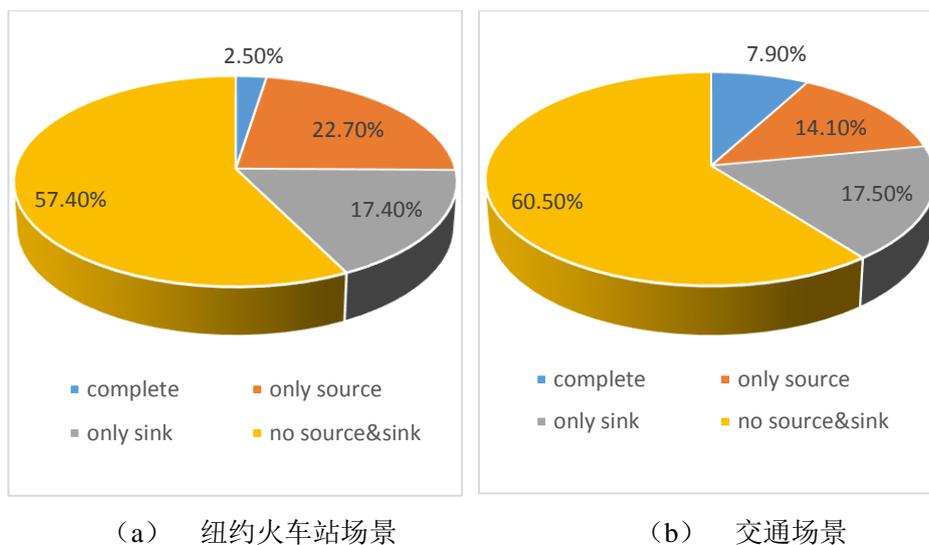


图 1-4 轨迹完整性统计图

综上所述,在当前社会背景下,智能视频监控技术具有广泛的研究和实用价值。运动目标轨迹分类作为集群场景智能监控技术实现的重要方法,不仅具有广泛应用的课题,同时是个具有挑战性的课题。

在本课题研究期间,本文作者参与并受到以下课题的资助:

1) “基于多源数据的飞行器进近威胁目标检测跟踪及行为预测”,“国家自然科学基金重点项目”(课题编号:61039003)

2) “多视角多姿态人体目标检测研究”,“国家自然科学基金重点项目”(课题编号:61271433)

1.2 研究进展及存在的问题

运动目标轨迹分类作为运动目标理解的重要研究内容之一,是实现智能监控的重要研究课题。本课题主要涉及监控场景中视觉轨迹分析、集群场景的轨迹分析和机器学习等领域,本章节主要针对相关领域的研究现状进行讨论。

监控场景根据其运动人流密度分为集群场景(Crowded Scene)和非集群场景(Sparse Scene)。在集群场景中,由于运动人流密度较大,所有运动目标的运动事件同时空发生、运动目标之间交互行为多见、运动情况复杂多变,被监控的运动目标之间遮挡频繁,现有的跟踪识别算法难以获取到一个运动目标的完整运动轨迹,获取到的轨迹片断性高;对于非集群场景,人流密度较小,运动情况简单,运动目标之间遮挡现象少见,基于视觉的跟踪识别算法能够获取到运动目标较为完整的运动轨迹。

轨迹数据获取作为轨迹分析的前提,主要是通过识别、跟踪算法提取图像序列中运动目标的轨迹信息。一般地,连续 T 帧运动目标的运动轨迹能够表示为一个流向量序列(Flow Vector Sequence)^[7],见公式(1-1)。

$$O_T = \{o_1, o_2, \dots, o_T\} \quad (1-1)$$

其中 o_t 表示该运动目标 t 时刻的运动状态,可以是运动目标的位置、速度、加速度等特征。通常的,跟踪识别算法直接捕获运动目标的位置信息,速度、加速度是在运动目标位置计算得到的。但是,这样的计算结果存在较大的噪音。因

此,运动轨迹分析通常采用 n 个时序相关的坐标点构成的流向量作为运动目标轨迹底层的特征表达, 见公式 1-2。

$$O = \{(x_1, y_1, t_1), (x_2, y_2, t_2), \dots, (x_n, y_n, t_n)\} \quad (1-2)$$

其中, 一个向量流 (Flow Vector) ^[7] (x_i, y_i, t_i) 表示 t_i 时刻运动目标的坐标位置是 (x, y) 。

1.2.1 非集群场景中运动目标轨迹分析

国内外学者对非集群监控场景中的运动轨迹分析已经进行了大量的研究。一般流程是通过构造距离测度算子对目标轨迹进行相似度衡量, 基于相似度衡量结果设计聚类、分类算法或者建模获得监控场景中运动目标的运动路径或者类别标号。由于运动目标速度各不相同, 识别跟踪算法获取到的运动目标轨迹长度往往不同。一条运动轨迹表示为一个向量流序列, 原始轨迹的长度不同, 因而表示为包含不等数量向量流的向量流序列。为了实现要求相同轨迹长度作为输入的相似度测度算子的相似度衡量, 一些轨迹分析工作需要对原始轨迹进行预处理, 保证其长度相同。因此, 非集群场景中运动目标轨迹分析一般流程见图 1-5。

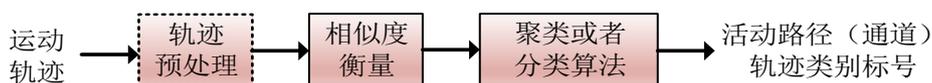


图 1-5 非集群场景中轨迹分析流程图

1.2.1.1 轨迹预处理

由于监控场景中光照的变化、遮挡以及获取的运动目标轨迹长短不一等原因, 原始的目标轨迹不能直接用来分析。为了高效的利用原始轨迹数据, 轨迹预处理是轨迹分析前的准备工作。轨迹预处理的方法主要包括: 轨迹标准化 (Normalization) 和轨迹降维 (Dimensionality Reducion) ^[2,7]。

1) 轨迹标准化

轨迹标准化的思想与统计学中归一化的思想类似, 目的是不同长度的原始轨迹表示成包含相同数量向量流的向量流序列^[2]。目前最常见的轨迹标准化的方

法是基于线性插值的重采样^[7-9]。在文献[8]中，Hu 等研究者提出用 FCM (Fuzzy C-means, 模糊 C 均值) 聚类算法实现运动轨迹基于时空的聚类 (Spatial-Based Clustering)，该聚类算法要求等长的目标轨迹数据作为输入。为了保证聚类算法的有效性，Hu 等研究者提出用基于更大时间间隔的重采样技术去除相邻两个重采样点之间的原始轨迹点后，采用线性插值方法实现输入聚类算法的所有轨迹具有相同长度。文献[9]采用同样的方法实现对原始轨迹的预处理，作为 FCM 聚类算法的输入。一般的，以最长的轨迹长度作为基准，进行线性插值，这种轨迹预处理的方法导致轨迹的空间过大，计算复杂度高。

2) 轨迹降维

轨迹降维实际上是变换轨迹的表达空间，将原始轨迹变换到方便分析和运算的低维空间^[2,7]。主要的轨迹降维技术包括 PCA (Principal Component Analysis, 主成分分析)^[2,7,10,11]、向量量化 (Vector Quantization) 技术^[2,5,7,12]。

PCA 是一种多元统计分析方法，通过线性变换降低数据集的维度，同时能够保持能量集中的特征，消除特征之间的冗余关联，低维空间中的特征向量之间保持正交^[13,14]。在文献[10,11]工作中，采用 PCA 方法将原始数据用低维空间的特征向量重新表达，提升聚类算法的效果同时降低聚类过程中的计算复杂度。

向量量化技术是将原始轨迹向量流序列通过量化转换为带有语义 (Semantic) 信息的字符序列。在文献[5,12]工作中，Wang 等研究者提出采用轨迹位置和速度方向量化技术将视觉轨迹特征转换成具有语义 (Semantic) 信息的字符项序列，作为分层贝叶斯模型 (Hierarchical Bayesian Models) 的有效输入，同时新的描述轨迹方法有效的降低了原始轨迹空间的维度。

1.2.2.2 相似度衡量

相似度衡量算法是计算轨迹距离的常用方法，主要包括欧氏距离 (Euclidean Distance)^[15-18]、DTW (Dynamic Time Warping Distance, 动态时间整合距离)^[15,17,19]、LCSS (Longest Common Subsequence, 最长公共子序列)^[15,17,20]、霍斯多夫距离 (Hausdorff Distance)^[15,17,21,22]，其中欧式距离算法要求等长的轨迹数

据作为输入。为了方便表达，本小节中轨迹数据用二维的向量流表示，见公式(1-3)，其中，第*i*条轨迹*n*时刻的位置是 (x_n^i, y_n^i) 。

$$T_i = \{(x_1^i, y_1^i), (x_2^i, y_2^i), \dots, (x_n^i, y_n^i)\} \quad (1-3)$$

1) 欧式距离

欧式距离是最常用的衡量两条相同长度轨迹之间的距离的算子。轨迹*i*和轨迹*j*的欧式距离 D_E 计算公式见(1-4)。

$$D_E(T_i, T_j) = \frac{1}{N} \sum_{n=1}^N \sqrt{[(x_n^i - x_n^j)^2 + (y_n^i - y_n^j)^2]} \quad (1-4)$$

欧氏距离常常和 PCA 配合使用^[15]。在距离衡量之前，用主成分分析对原始轨迹进行空间变换，有效的降低原始轨迹空间，在新的轨迹空间中，轨迹具备低纬度和相同长度的特点。两种算法有效的结合实现了用有效的特征表达轨迹的基础上去掉噪音和冗余的特征，提升聚类分类等后续工作的效果。

2) DTW 距离

DTW 最初是基于动态规划 (Dynamic Programming, DP) 的思想解决语音识别领域发音长短不一的模版匹配问题^[23]。现在大量工作用 DTW 寻找两条轨迹最小代价的规整路径 W ，动态时间规整距离 D_{DTW} 见公式(1-5)，其中， w_k^{ij} 是轨迹*i*与轨迹*j*的规整路径 W 的第*k*项。

$$D_{DTW}(T_i, T_j) = \min \left\{ \frac{1}{K} \sqrt{\left[\sum_{k=1}^K w_k^{ij} \right]} \right\} \quad (1-5)$$

3) LCSS 距离

文献[20]用最长公共子序列算法作为不同轨迹的相似度衡量算子，最长公共子序列进行轨迹匹配时允许个别的轨迹点不进行匹配，然而欧氏距离和动态时间规整距离测度算法将所有的轨迹点都进行匹配，这种特点能够有效避免原始轨迹数据中的噪音点和异常点，具有更好的鲁棒性。

4) 霍斯多夫距离 (Hausdorff Distance)

霍斯多夫距离用于衡量两个点集间的距离，用于轨迹衡量时将轨迹看作无序的点集，忽略其空间时序以及方向的关系。文献[21]中提出用 4D 的矢量流序列作为轨迹底层的表达方式，在位置信息的基础上加入了速度信息，用多维的霍斯多夫距离进行相似度衡量。

1. 2. 2. 3. 轨迹聚类和分类 (Trajectory Clustering and Classification)

轨迹聚类和分类是轨迹分析常用的方法。轨迹聚类作为非监督学习的代表，主要是基于轨迹相似度对轨迹之间关系的衡量，将相似度高的轨迹聚集在一起，而这些相似的轨迹簇就形成一个活动路径（通道）；轨迹分类作为监督学习的代表，通过分类算法对具有类别标定的训练数据的学习和分析，获得训练集中每条轨迹的类别标号。

目前，常用的轨迹分析算法包括迭代优化 (Iterative Optimization) [7,18,24]、语义聚类 (Semantic Clustering) [5,12,25-27]等。

1) 迭代优化

K-means (K 均值) [24]和 FCM[18]是两个经典的基于距离测度算子的迭代优化的聚类算法，是以某种距离测度算子作为迭代优化的目标函数，不断迭代调整聚类中心，直至属于每个聚类中心的样本不再改变迭代结束。在 K 均值聚类算法中，每个样本只能属于所有类别中的一类，这样会带来一些问题：如果某个样本到每个聚类中心的距离都是一样的，K 均值无法处理。因此，从根本上讲，K 均值算法无法区分所有样本都属于同一个类的情况。在模糊 C 均值算法中，样本可能属于任何一类，但是，属于不同聚类中心的可能性是不同的。一般的，欧氏距离是迭代优化中常用的距离测度函数，因此，K-means 和 FCM 聚类算法要求输入的轨迹数据长度相同。文献[18]采用线性差值对原始轨迹数据进行预处理，文献[25]将 PCA 与聚类算法配合使用。

2) 语义聚类

语义聚类方法主要源自于自然语言分析的研究工作，在轨迹的数据分析中引入了文档-主题的聚类思想，认为在轨迹的底层和高层特征表达之间存在基于

语义的中层特征，即主题。在文献[4,12,25-27]中，将自然语言分析领域的主题模型（Topic Model）引入计算机视觉领域进行轨迹聚类分析。在这些研究中，轨迹数据被视为 BOW（Bag-of-Words，词包），轨迹点经过的区域经过量化处理视作词项（Word），轨迹被视作由一系列无序词项构成的文档（Document），整个轨迹数据集被视作由一系列文档构成的语料库（Corpus）。基于语义聚类的思想认为在底层的词项空间（Word-Space）和高层的文档空间（Document-Space）存在一个基于语义的中层特征空间，叫做主题（Topic）。这里的主题类似与 K-means 和 FCM 聚类算法中的聚类中心。基于语义的聚类方法并不需要统一轨迹长度。

1.2.2 基于视觉的集群目标轨迹分析

如本章第二节所述，监控场景按照人流密度分为集群场景和非集群场景。

非集群监控场景中运动目标较为稀疏，运动目标在时间序列上出现的较为单一，运动目标间相互遮挡现象少见，基于计算机视觉的跟踪识别算法能够捕获到较为完整的运动目标的运动轨迹，完整的运动轨迹在形状、方向等特征上具有类别可分性，基于距离测度算子的聚类方法能够对监控场景的运动轨迹进行有效的分析。

集群监控场景中运动目标较为密集，监控场景中所有的运动目标的运动行为几乎同时发生，运动目标之间交互行为多见，运动目标之间相互遮挡严重，基于计算机视觉的跟踪识别算法捕获到的运动轨迹片断性高。

文献[9]中将监控场景的进口区域（Entry zones）和出口区域（Exit zones）称为一类 POI（Point of Interest，兴趣点）。POI 是指兴趣的事件发生的区域^[7]，主要指监控场景中运动目标突然出现、消失、停留以及聚集的区域，如监控场景的进出口区域是运动目标常常出现和消失的地方；火车站场景中的售票点是运动目标常常停留和聚集的区域。如图 1-6 所示，椭圆区域是该监控场景的 POI，若运动轨迹的始发点和结束点都在兴趣点区域，则称轨迹是完整的，可用示意图 1-6 中黄线表示；反之，则是片断轨迹。集群场景中运动轨迹的高度片断导致无法延续基于距离测度算子的聚类分类方法实现轨迹分析。为了提高轨迹聚类的效

果，非集群场景的轨迹分析工作中少量的不完整轨迹数据被当作噪音数据，不参与轨迹分析^[9]。如本章第一节对两个公开数据集的轨迹进行统计，不完整轨迹占整个数据集的 90% 以上，因此，无法通过传统方法实现集群场景的轨迹聚类。

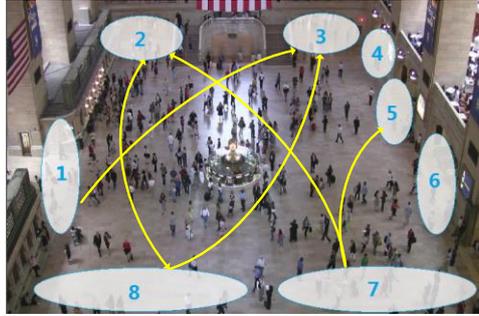


图 1-6 监控场景中 POI 区域示意图

集群场景按照其场景结构可分为有结构（Structured Crowded Scene）（图 1-7（a））和无结构（Unstructured Crowded Scene）（图 1-7（b））^[2,28]。在有结构的监控场景中，存在能够约束人群运动的道路、障碍物等交通设施或者物体等，运动目标有特定的运动，常见的有交通场景，十字路口、马拉松跑道等。图 1-7（a）中，所有的人群都绕着 U 型的马拉松跑道逆时针运动。这类监控场景中，虽然拥挤，但是运动行为单一，运动方向和运动模式通常固定不变。在无结构的监控场景中，通常不存在能够约束人群运动的交通设施和物体，运动目标没有特定的运动规则，常见的有无障碍广场、大厅等。图 1-7（b）中，除了火车站中心的花坛之外没有其他能够约束人群运动路线的设施和物体。这类监控场景中，人群运动随机，方向多变，支持一种或多种运动形态。



(a) 有结构集群场景



(b) 无结构集群场景

图 1-7 有结构集群场景和无结构集群场景

文献[29]提出，自然界的生物系统中存在一种自组织（Self-organized）的集群行为，这种自组织过程（Self-organized Process）被定义为系统自发的出现更大

规模的结构,这种结构超越系统子单元的局部的相互作用。新的结构能够体现了一种新的特征,并不是系统子单元特性的简单叠加。自组织过程让整个系统呈现了一种新的特性。文献[30]提出一个集群环境表面上看起来混乱无序,实际上有其固有的集群的动态特性。处于同一个集群环境的不同个体,尽管相互没有关系,在各自运动的过程中无形的受到集群中其他个体运动的影响而调整自己的运动模式,直至整个系统趋于一种稳定的状态,此时,系统会呈现出新的集群特性。如图 1-8 (a) 所示^[29],个体鱼之间相互协调让整个群体的运动变得有序,系统趋于稳定状态时呈现出一个环形的旋窝;图 1-8 (b) ^[29]中,蚂蚁从一个方形的环路的一端去往另一端,个体的道路选择受到其他个体的影响,集群都聚集在两条道路的其中一条上。将这个理论具体到人群系统中,互不相识的个体处于一个集群环境中具有相同的运动和目的地,个体在进行各自的运动的过程中,受到其他个体的影响,相互调整自己的运动,最终集群系统以固定的模式稳定运动。如图 1-8 (c) ^[29],方向不同的个体受到周围和对面个体运动的影响不断调整自己的运动,最终形成方向相反的两种运动模式。从社会心理学的角度来讲,任何个体在任何一个环境中都想以最捷径的方式达到自己的目的。比如无障碍环境中,个体总是愿意沿着直线运动。在集群的环境中,拥有相同目的的个体选择了一条最捷径的方式开始运动,运动发生的过程中与环境中其他个体相遇,为了避免碰撞,个体依据其他个体相互调整自己运动过程,最终达到一种稳定状态。换一种说法就是随着时间的推移,个体的运动状态会渐渐弱化,取而代之的是一种“从众”的运动模式,这种运动模式具有自发性和无组织性。

近些年,Wang 等学者将自然语言处理领域的相关技术引入计算机视觉领域,最典型的是主题模型。自然语言分析领域,主题模型是挖掘文档主题的常用技术,将其用在计算机视觉领域,能够有效的挖掘集群运动模式,并将其用于广泛的应用中,如聚类^[4,25-27]、分类、跟踪^[28]等。通常的, BOW 转换是这类研究工作的基本步骤。所谓 BOW 模型就是基于字典的建模方法,每一个文档被视为一个包,包含了来自字典的词,文档中的词被忽略了次序。BOW 的建模思想被广泛的应

用在计算机视觉中，在图像处理中，一张图片被视作一个文档，从图像中提取出来的特征被视作词项，同样的思想应用在监控视频处理中。

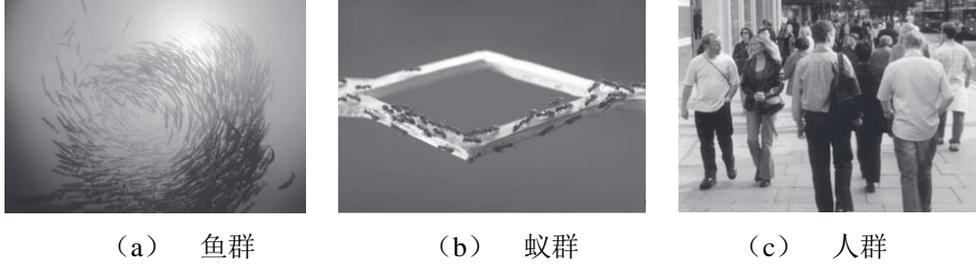


图 1-8 不同集群的自组织行为

文献[28]中，监控视频被截成时间上无重叠的视频段，每个视频段被视为一个文档，底层的视觉(运动轨迹的位置和速度)特征被量化后视作 BOW 中的词，视频段被认为是由一系列集群行为构成，这里的集群行为就是 CTM (Correlated Topic Model, 关联主题模型) 学习的主题，学习到的集群行为将作为非结构集群场景中跟踪工作的先验信息。

人流密度极大的集群场景，基于视觉的跟踪识别算法难以获取个体完整运动轨迹，高片段的轨迹在形状和方向上并不具有区分性，但是，片段轨迹在时空上具有一定的联系。文献[5]提出 RFTM (Random Field Topic Model, 随机条件场主题模型) 通过对片段轨迹进行建模，学习集群场景中的运动模式。具体的讲是用 MRF (Markov Random Field, 马尔科夫随机场模型) 寻找不同轨迹之间时空关联，这种关联被当作主题模型的先验信息进行集群场景中运动的模式学习。在该研究工作中，轨迹被视作文档，底层的视觉特征(运动轨迹的位置和速度)特征经过量化被视作 BOW 中的词项，主题模型用来挖掘场景中语义运动模式。

文献[25,26]在文献[4]工作的基础上，提出 BCTM (Belief based Correlated Topic Model, 基于置信的关联主题模型) 学习场景运动模式。与文献[4]不同的是，用 CTM 作为运动模式学习的模型，避免了文献[5]中 LDA (Latent Dirichlet Allocation, 隐狄利克雷分布) 模型假设数据独立带来的缺陷。文献[26]提出利用 CTM 中高斯分布的协方差刻画主题之间的关联，通过关联有效的解决了模型学习中层特征的过分割和未完全分割的问题，有效提高了聚类的效果。

1.3 本文主要研究内容

本课题研究的目的是要在对集群监控场景运动目标轨迹分析技术的基础上,结合轨迹分析在轨迹预处理合理性、轨迹表达准确性、运动模式学习的有效性及其应用普遍存在的问题,提出集群监控场景运动目标轨迹分类的解决方案,并在此方案的基础上进行更细致的优化和改进。具体来说,就是要实现输入为监控视频输出为每条未知轨迹的标号,无人工干预的智能监控方案,并能够达到准确性、实用性、稳定性等要求。为了实现本课题研究的目标,按照模型的判别式扩展和生成式扩展两种思想,本课题的任务可以分为两个模块,分别是基于 BCTM 与多类 SVM 的集群目标轨迹分类和基于 Multi-CTM 集群目标轨迹分类,研究工作的整体框架如图 1-9 所示。

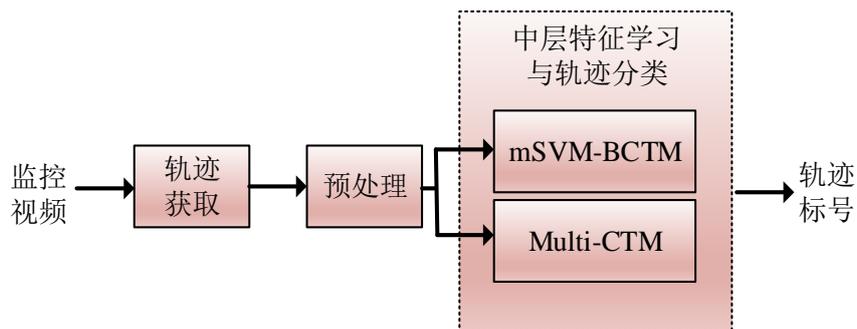


图 1-9 本文整体框架图

1. 基于 BCTM 与多类 SVM 的集群目标轨迹分类

在集群场景轨迹分类解决方案的设计与实现的研究中,包括详细任务如下:

- (1) 结合需求的特点,对同类轨迹分析进行广泛的研究,分析现有运动轨迹获取的算法,选择适合本文需求的轨迹提取算法;
- (2) 针对集群场景中运动目标相互遮挡频繁的难点问题,在现有的轨迹预处理的基础上,改进并提出适合本文研究目标的轨迹预处理方式;
- (3) 在对原始运动轨迹数据预处理的基础上,对主题模型进行广泛研究和深入学习,采用适当的模型进行集群行为的模式学习;
- (4) 根据学习到的行为模式,设计合理的分类流程,实现运动目标轨迹分类。

2. 基于 Multi-CTM 的集群目标轨迹分类

在基于主题模型的集群轨迹分类的解决框架实现的基础上，为了简化和优化框架，提高分类的准确率，设计和实现基于多关联主题模型的集群目标轨迹分类框架，包括详细任务如下：

(1) 基于对主题模型及其应用的广泛研究和学习，提出 Multi-CTM，有效的实现集群场景中运动模式和未知轨迹类别标号的同时输出；

(2) 设计并实现弱监督分类学习过程，有效的减少了监督学习中人工标定带来的代价。

根据以上课题任务，本论文的组织结构如下：

第一章 绪论。本章介绍本论文研究的问题背景和意义、当前研究领域的研究状况的综述、论文的研究任务以及组织结构。

第二章 相关技术研究。本章是承上启下的章节，主要是对使用的相关技术的介绍，包括运动轨迹获取算法、轨迹预处理算法、主题模型和分类算法等内容。

第三章 基于 BCTM 与多类 SVM 集群目标轨迹分类。本章主要介绍轨迹预处理方法的改进、主题模型的选择、分类流程的设计与实现以及相关的实验验证结果。

第四章 基于 Multi-CTM 集群轨迹分类。本章主要介绍多关联主题模型的提出和实现、若监督分类过程的设计和实现以及相关实验验证结果。

第五章 结论与展望。本章主要对本文的研究工作进行总结以及对下一步工作的开展进行展望。

第二章 相关技术研究

前一章主要简要介绍本论文的问题背景和意义、当前相关领域的研究综述、本文的主要研究内容和本文结构安排。本章是第一章的延伸，将介绍第三章、第四章研究工作中使用到的算法和相关技术。

2.1 视觉跟踪算法

跟踪算法是自动获取监控视频中运动目标轨迹的关键方法，KLT^[4]能够快速地从图像序列中获取多个目标的运动轨迹。

KLT 是典型的基于光流的跟踪算法，满足光流法的三个假定：

- ① 相邻两帧之间亮度恒定；
- ② 相邻视频帧的相邻两帧时间相同或相邻两帧物体的运动是“小运动”；
- ③ 保持空间一致，临近点运动相似，保持一致。

基于光流法的目标跟踪的基本原理是对一个连续的视频帧序列，对每一帧的视频，利用一定的目标检测方法，检测到可能出现的目标后，用其最具有代表性点作为关键特征点（通常是角点），对于之后的视频帧，寻找上一帧关键特征点在当前帧出现的位置，反复迭代进行，即可实现目标跟踪。

用 (x, y, t) 表示 t 帧某点在图像中的位置为 (x, y) ，根据亮度恒定（假设①）和微小运动（假设②）两条假设的基础上，可以得到公式 2-1，

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (2-1)$$

其中， dx ， dy ， dt 分别是 x 方向， y 方向以及时间 t 上的偏移量。根据空间一致性（假设③）的假设，对于同一个窗口 W 中的所有 (x, y) 具有相同的运动，即同时往一个方向偏移了 (dx, dy) ，因此，寻求相邻两帧图片关键点匹配问题即可转换为寻求公式 2-2 的最小值的优化问题，其中， $\varepsilon(d)$ 可视为误差函数。

$$\varepsilon(d) = \varepsilon(dx, dy) = \sum_W (I(x, y) - I(x + dx, y + dy, t + dt))^2 \quad (2-2)$$

2.2 MRF

基于前面的论述，集群场景中，虽然跟踪识别算法获取到的目标轨迹非常短，片断性高并且混乱，但是，这种片断性的轨迹在时间和空间上具有一定的关联。如图 2-1 所示，图中线段表示通过跟踪算法获取的运动轨迹，箭头表示轨迹的方向，其中紫色线段表示完整轨迹，红色线段表示丢失了后半段的片断轨迹，可称为有进口的轨迹，黄色表示丢失了前半段的片断轨迹，可称为有出口的轨迹，绿色表示丢失了前半段和后半段的片段轨迹，可称为没有进口也没有出口的轨迹。将满足时空一致性关系的片断轨迹连接成完整轨迹的过程与生成树的过程类似，树根由有进口的轨迹构成（如图 2-1 中红色轨迹），树叶由有出口的轨迹（如图 2-1 中黄色轨迹），从树根到树叶的路径被视为一条完整的轨迹，构成完整轨迹的片段轨迹保持时空一致性，一棵树包含一条或者多条完整轨迹。特别的，完整轨迹中的片段轨迹并不一定是同一个运动目标的轨迹，但是能够反映集群场景中的运动模式。文献[5]研究工作中提出的 MRF 就是基于上述原理捕捉集群场景中片段轨迹之间时空关联。

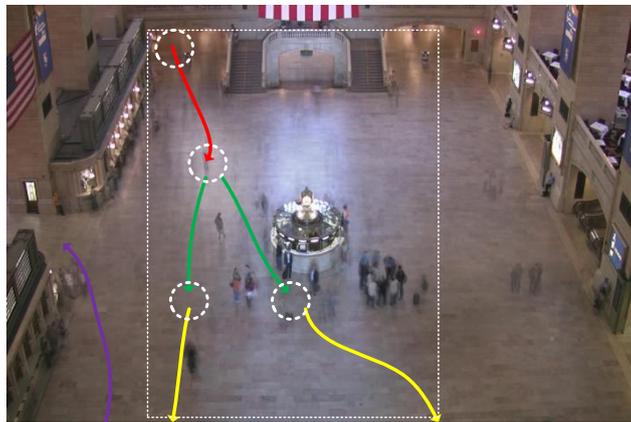


图 2-1 片段轨迹时空一致性示意图

为了方便阅读和查阅，本文仍然延续文献[5]中的定义。轨迹 p 开始于 t_p^s 时刻，结束于 t_p^e 时刻。它在起始点 (x_p^s, y_p^s) 和终止点 (x_p^e, y_p^e) 的速度表示为 $v_p^s = (v_{px}^s, v_{py}^s)$ 和 $v_p^e = (v_{px}^e, v_{py}^e)$ 。公式 (2-3) 定义了轨迹之间的邻域关系，具体的

讲，公式 (2-3) 中 *I* 定义了成为邻域的轨迹要保证时间一致性，轨迹 *q* 的起始点出现的时间帧 t_q^s 满足轨迹 *p* 终止点出现时间帧 t_p^e 的 *T* 时间内；公式 (2-3) 中 *II* 定义了成为邻域的片段轨迹要保证空间一致性，轨迹 *q* 的起始点位置 (x_q^s, y_q^s) 要满足在轨迹 *p* 的终止点 (x_p^e, y_p^e) 的 *S* 范围内；公式 (2-3) 中 *III* 定义了成为邻域的轨迹要保证速度方向一致性，轨迹 *q* 的起始点速度 (v_{qx}^s, v_{qy}^s) 方向与轨迹 *p* 终止点的速度 (v_{px}^e, v_{py}^e) 方向保持一致，*T*、*S*、*C* 控制轨迹 *p* 和轨迹 *q* 时间、空间、速度方向一致性的参数。

$$\begin{aligned}
 I. \quad & t_p^e \leq t_q^s < t_p^e + T \\
 II. \quad & |x_p^e - x_q^s| + |y_p^e - y_q^s| < S \\
 III. \quad & \frac{v_p^e \cdot v_q^s}{\|v_p^e\| \cdot \|v_q^s\|} > C
 \end{aligned} \tag{2-3}$$

$$W(q, p) = \begin{cases} \exp\left[\left(\frac{v_q^s \cdot v_p^e}{\|v_p^e\| \cdot \|v_q^s\|} - 1\right)\right], & \text{if (2-3) satisfies,} \\ 0, & \text{otherwise.} \end{cases} \tag{2-4}$$

如果轨迹 *p* 和 *q* 满足公式 (2-3) 中 *I*、*II* 和 *III*，称轨迹 *q* 是轨迹 *p* 的邻域轨迹，轨迹 *q* 和轨迹 *p* 具有时空一致性，反之，若不满足其中一项，则称轨迹 *q* 和轨迹 *p* 不具有时空一致性关联。当轨迹 *p* 和轨迹 *q* 满足时空一致性，此时，轨迹 *q* 与轨迹 *p* 首尾相连的能量通过公式 (2-4) 计算，可以看出，当轨迹 *q* 的起始点 (x_q^s, y_q^s) 和轨迹 *p* 的终止点 (x_p^e, y_p^e) 速度方向一致程度越高，能量函数 $W(q, p)$ 越大。

MRF 具有马尔科夫随机过程的原型，具体来讲，某条轨迹是否能够连接到这条轨迹树上只与前一条轨迹有关，与其他轨迹无关，连接能量函数可视为马尔科夫随机过程中的转移概率，能量越大，就越能表现同一个运动模式。

2.3 BOW

2.3.1 文本 BOW 字典构建

BOW 模型起源于自然语言处理和信息检索领域，该模型忽略掉文本的语法和次序，用一组无序的单词来表达一段文字或者一个文档。例如，在 BOW 模型中，“a red apple”与“red a apple”等价。简单描述一下自然语言处理领域“词包”模型构建字典的过程，下面是两个简单的文本文档：

D1: “Bob likes to play footb. John likes to play basketball.”

D2: “Bob likes to watch TV. Lily likes too.”

对这两个简单的文本文档建立“词包”模型，用索引来表示文档中出现的单词，基于 C 语言的习惯，索引从 0 开始：

表 1-1 语料库中单词与索引对应表

单词	Bob	likes	to	play	football	John	basketball	watch	TV	Lily	too
索引	0	1	2	3	4	5	6	7	8	9	10

由两个文档构成的语料库中包含 11 个不同的单词，构建了如表 2-1 的“词包”字典。用字典对两个文档进行表示，两种常用的表示方式如下：

1) D1: [1,2,2,2,1,1,1,0,0,0,0]

D2: [1,2,1,0,0,0,0,1,1,1,1]

2) D1: [(0:1),(1:2),(2:2),(3:2),(4:1),(5:1),(6:1)]

D2: [(0:1),(1:2),(2:1),(7:1),(8:1),(9:1),(10:1)]

表示法 1) 类似于直方图表示法，向量的每一维数值代表了相应单词在文档中出现的次数，采用此表示法表示的所有文档的维度与字典维度相同；表示法 2) 中每一维括号代表（单词索引：个数），用此表示法表示的文档维度不一定与字典维度相同，不同文档的维度也可能不同，没有在文档中出现的单词索引不会在表达式中出现。这样的表示法并不能保证单词在原文档中的顺序。

2.3.2 图像 BOW 字典构建

词包模型可以认为是一种统计直方图，Feifei Li 在[31]中提出用 BOW 表达图像的方法。在图像的 BOW 表达中，每一幅图片视作一个文档，具体构建字典过程如下：

1) 对每一幅图像进行块 (patch) 分割，一副图像 patch 的个数见公式 (2-5)，其中 $width * height$ 表示图像像素， $p * p$ 表示 patch 的像素大小， $step$ 为步长。例如，图 2-2 中，假设图像像素为 $256 * 256$ ，一个 patch 的像素大小为 $16 * 16$ (红色方框)， $step$ 为 16，每次向右或者向下移动一个步长形成另一个 patch，图 2-2 的图像中一共形成了 256 个 patch (图中白色虚线小方框)；

$$Number\ of\ Patches = \left(\frac{width - p}{step} + 1\right) * \left(\frac{height - p}{step} + 1\right) \quad (2-5)$$

2) Patch 和图像的视觉特征表达。通常的，采用 SIFT (Scale-invariant feature transform, 尺度不变特征转换) 算子对每个 patch 进行特征提取，每个 patch 由 128 维 SIFT 特征向量表达，整幅图像由 N 个的 128 维 SIFT 特征表达， N 为图像中 patch 的个数，图 2-2 中矩阵的每一行表示一个 patch 的 SIFT 特征；

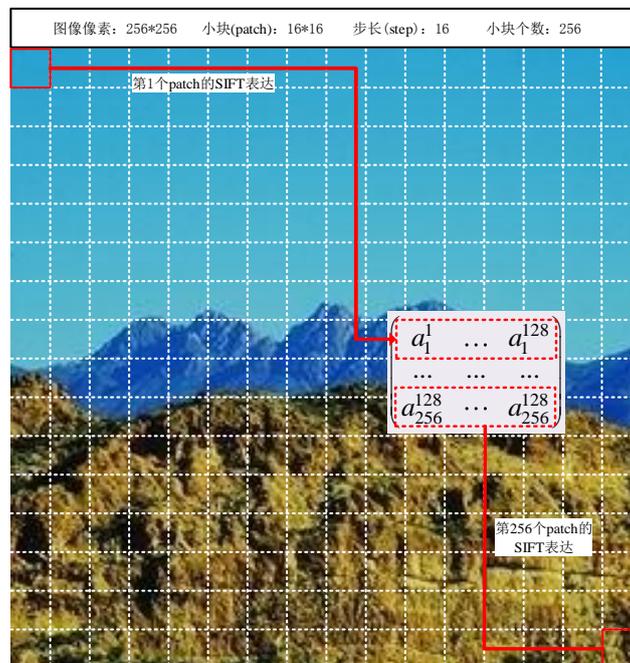


图 2-2 一副图像 SIFT 特征表达示意图

3) 字典或码本 (Codebook) 生成。对所有图像的所有块的 SIFT 特征进行聚类, 通常用 K-means 聚类算法, K 个聚类中心即构成字典中的词项。

4) 图像 BOW 表达。图像的每一个 patch 的 SIFT 特征寻找距离最近的聚类中心进行表达, 并进行直方图统计。

2.3.3 轨迹表达中 BOW 字典构建

近些年, 用词包模型对运动轨迹进行特征表达是非常常见的方法, 文献[5,25-27]中, 轨迹被视作文档, 量化的轨迹点被视作文词, 构建字典过程如下:

1) 通过对轨迹底层特征进行量化构建码本。具体的讲, 图 2-3 中, 像素为 720*480 的视频序列被分割为 10*10 的单元, 共有 72*48 个单元, 速度方向量化为 4 个 (图 2-2 图像上方黄色圆环), 每个单元 (Cell) 中同方向的所有像素点表示相同的单词 (Word), 因此, 构建 BOW 字典维度为 72*48*4。

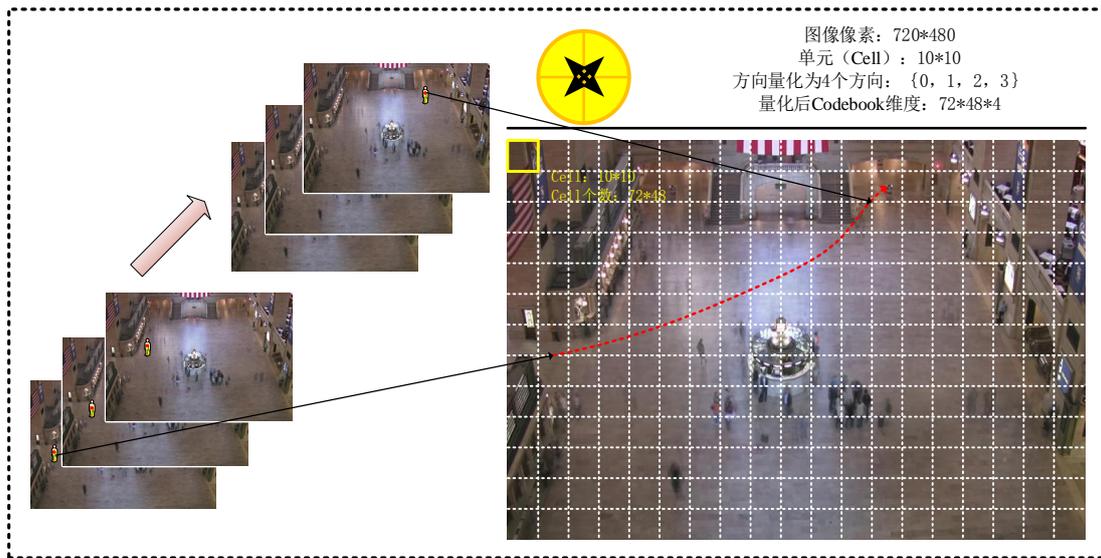


图 2-3 轨迹量化示意图

2) 轨迹 BOW 表达。轨迹的每个点进行“词项”转换后, 对每条轨迹中的词项进行直方图统计, 轨迹点 (x, y) 的“词项”转换规则见公式 (2-6), 其中, $width*height$ 是视频序列像素, $direction$ 是轨迹点 (x, y) 量化后的方向, 其值属于集合 $\{1, 2, 3, 4\}$, C 是分割单元的像素大小。

$$word = (x / C) * (height / C) + (y / C) + direction * (width / C) * (height / C) \quad (2-6)$$

2.4 主题模型

传统的衡量两篇文档的相似度的算法是基于统计两篇文档中共同出现的单词的个数。但是，这样的方法是不精确的。两个文档的相似程度不仅取决于两个文档中共同出现单词的个数，还取决于两个文档文字背后语义上的关联程度。例如，有两段文字，分别如下：

- 1) 手机更新换代非常快。
- 2) 小米又降价了。

两段文字中没有重复的单词，如果采用统计两段文字中共同出现单词个数的方法判断两段文字肯定不相似，但是，如果交给人脑来判断，这两段文字一定是相似程度非常大。因为，第一段文字中有“手机”，虽然第二个句子中“小米”可能是指谷类，但是人脑非常自然的会把“小米”关联到小米公司的手机产品上。这种文字背后的语义关联对自动进行文本分析、搜索引擎算法的研究非常重要。主题模型是用于挖掘文字背后隐含的主题。所谓“主题”，就是一篇文档或者一段话所表达的中心思想。

主题模型中有 4 个重要的概念，分别是语料库(Corpus)、文档(Document)、单词(Words)、主题(Topic)。一个文档由多个单词构成，多个文档构成一个语料库，每个文档都有其要表达的中心思想，即主题，同时，单词能够表达不同的主题，在不同的语义环境中，能够表达不同的主题。例如，“小米”既能表达谷类食品，也能够表达小米手机。

创作一篇文章，首先应该确定文章的主题，然后在该主题下选择能够表达主题的单词进行表达，假设该文章中包含 N 个单词，那么如此反复进行 N 次主题下的单词表达，即可构成一篇文章。实际上，主题模型是文章创作的一个逆过程，已知的是文档和文档中的单词，通过对这些单词和文档的建模学习文档表达的主题，有一点不同的是，在主题模型假设文档中的单词是无序的。

主题模型是基于概率的统计模型，一般步骤是先假设数据的概率分布，根据假设对数据进行建模，通过概率表达式迭代求解模型参数。最经典的主题模型是

LDA, 文献[32]在谷歌学术上的引用已经达到 11035 次。在最经典的主题模型之 N 前, 大致经历了 Unigram Model (一元模型)、Mixture of Unigrams (混合一元模型) 和 PLSA (Probabilistic Latent Semantic Analysis, 概率潜在语义分析)。

Unigram Model^[32]被认为是主题模型的起源, 实际上, 该模型中并没有主题的含义。模型认为一个文档由 N 个独立的单词构成, 因此, 要想形成一篇文档, 需要独立进行 N 次单词选择。在此过程中, 随机变量单词被假设服从多项分布。

Mixture of Unigrams Model^[32]认为每个文档具有一个主题, 文档由个独立的该主题下的单词构成, 因此, 文档的形成过程自然的分为两个步骤, 首先确立一个主题, 再在这个主题下独立进行 N 次单词选择。在此过程中, 文档的主题被定义为随机变量, 服从多项分布, 同样的, 随机变量单词被假设服从多项分布。

PLSA^[32,33]中, 文档被视为多个主题的混合, 每个单词都对应主题的概率分布。这样的假设更符合实际生活中的文档, 例如, 一篇文章可以同时表达医学和生物遗传两个主题。在这个主题模型中, 模型的参数随着语料库中文档个数的增加而线性增长, 造成大量的计算量; 另外, 模型中一个文档对应一个主题的分布参数, 容易出现过拟合现象。

基于以上对主题模型简单的铺垫, 下面着重介绍 LDA 和 CTM。

2.4.1 LDA

LDA^[32]是 Blei 在 2003 年提出, 该模型在 PLSA 模型的基础上, 将参数分布化, 有效的解决了 PLSA 模型中参数数量多、计算量大以及过拟合的问题, 成为了主题模型的经典。隐狄利克雷分布是本文研究工作的基础模型, 因此, 下面对 LDA 进行详细的介绍。

LDA 是分层贝叶斯模型, 第一层数据服从多项式分布, 第二层数据服从狄利克雷分布。对于语料库中的每一篇文档, 都对应一个主题分布, 对于每一个主题, 都对应一个单词分布。例如, 图 2-4 是 LDA 模型的原理示意图, 图中文档来自网页[34], 文档中不同颜色的词项表示不同主题, 比如“百度”和“框计算”之类的词项属于“百度”这个主题, “必应”和“微软”是主题“微软”的词

项。该文档包含 4 个主题，分别是“百度”，“微软”，“谷歌”和“市场”，每个主题对应一个词袋，词袋中的词项以一定的概率表达该主题，4 个主题对文档主题的表达有不同的贡献，被称为文档的主题分布，图中最右边的直方图即表示该文档的主题分布。

LDA 是一个典型的生成模型，定义的生成过程和人脑创造文章的过程非常相似。对于语料库中的每一篇文章，具体生成过程定义如下：

- 1) 首先确定文档包含的单词个数 N ；
- 2) 从主题分布中抽取一个主题；
- 3) 在 2) 中被抽到的主题对应的单词分布中抽取一个单词；
- 4) 重复步骤 2) 和步骤 3) N 次，直到文档中 N 个的单词都抽取完毕。

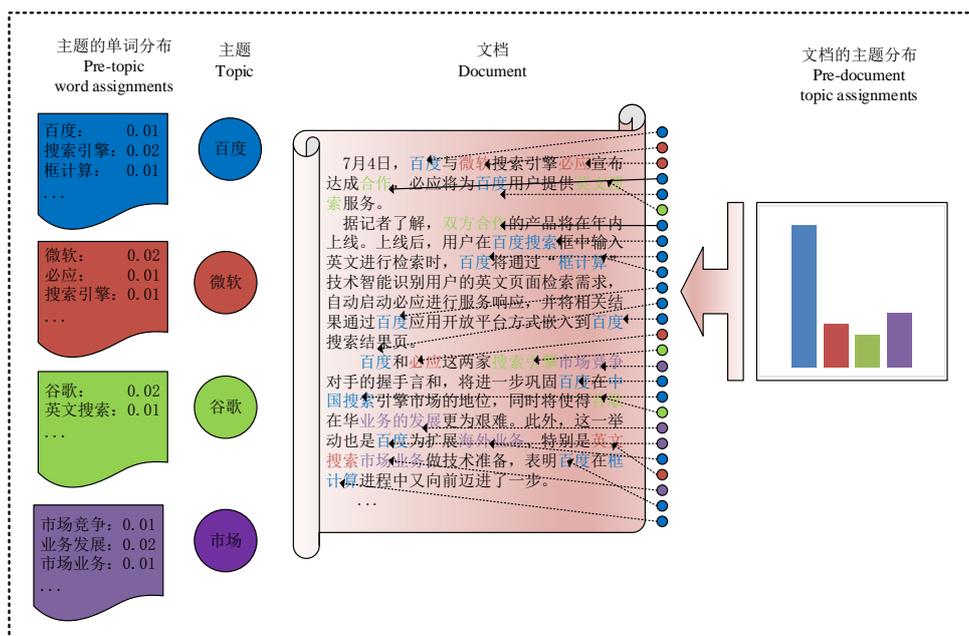


图 2-4 LDA 原理示意图

2.4.1.1 图模型

为了便于阅读和查阅，本文中仍然延续文献[32]中定义。模型的定义中，语料库由 M 个文档构成，文档由 N 个无序单词构成，语料库、文档的数学表达式分别见公式 (2-7) 中的 I 、 II 。

$$\begin{aligned} I. \text{ Corpus} &= (D_1, D_2, \dots, D_M) \\ II. D &= (w_1, w_2, \dots, w_N) \end{aligned} \quad (2-7)$$

LDA 的图模型见图 2-5 (a)，图中阴影圆圈表示已知变量，也称作可观测变量（如 w ），非阴影圆圈表示潜在变量（如 θ 、 z 等），没有圆圈的算子表示模型参数（如 α 、 β ），矩形方框表示重复的进行抽样， M 和 N 分别为抽样的次数，箭头代表两个变量的条件依赖性。具体的，参数 β 是一个 $K * V$ 的矩阵， K 表示主题的个数， V 表示单词字典的维度，字典中的单词来自语料库所有文档中不重复的单词，故有 β_{kv} 表示字典中第 v 个单词属于主题 k 的概率，矩阵一行表示一个主题特征的单词分布，矩阵的一列表示一个单词能够表达不同主题特征的概率分布。文档的主题分布变量 θ 服从参数为 α 的狄利克雷分布，即 $\theta \sim \text{Dir}(\alpha)$ ；模型的“隐”体现在隐变量 z 上，表示文字背后隐含的主题，在已知主题分布 θ 的条件下，对于文档中每一个单词的主题 z 服从参数为 θ 的多项式分布，即 $z_n \sim \text{Multinomial}(\theta)$ ；观测变量 w 表示单词在已知单词主题 z 和单词字典 β 的条件下，服从参数为 β_z 的多项分布。LDA 模型建立了一个“文档-主题-单词”的三层结构，同样的，模型的变量和参数也分为三个级别：

- 1) 语料级别变量-- α 和 β 称为超参数，整个语料库只有一个；
- 2) 文档级别参数-- θ ；语料库中每个文档都要进行一次主题分布变量 θ 的抽取，在 PLSA 模型中，主题分布是多项式分布的参数，LDA 模型将该参数分布化，避免了随着文档个数增加模型参数线性增长的问题，同时有效解决了 PLSA 模型中过拟合的问题；
- 3) 单词级别参数-- z ；每个单词都对应一个隐变量 z ，也就是说变量 z 抽取的次数和单词抽取的次数是一样的。

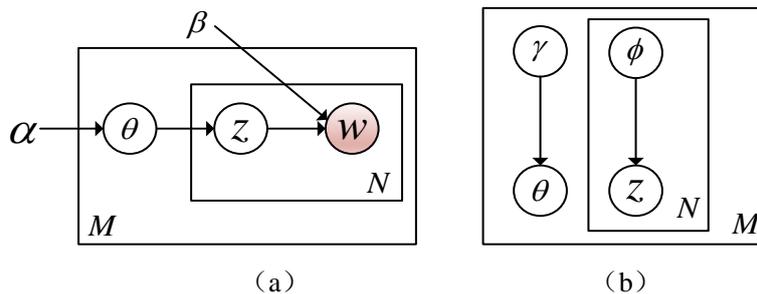


图 2-5 LDA 图模型 (a) 和近似图模型 (b)

2.4.1.2 模型公式推导

LDA 的参数推导大致有两种方法：一种是变分推断，另外一种为吉布斯采样。为了更契合本文的研究内容，本文仅介绍变分推断的方法。

对于一篇文档，随机变量 θ 、 Z 和 D 的联合概率见公式 (2-8)，其中 Z 是文档 D 中所有单词的主题集， z_n 表示文档 D 中第 n 个单词的主题。由于变量 Z 和 θ 不可观测，将其进行积分求和得到公式 (2-9)。公式 (2-9) 称作模型的似然函数，公式 (2-10) 称作模型的后验概率。模型的推导任务简单来讲就是求解出似然概率最大时候的参数值作为模型参数，参数进行后验概率推断后，得到隐变量的值。

$$p(\theta, Z, D | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta) \quad (2-8)$$

$$p(D | \alpha, \beta) = \int p(\theta | \alpha) \left(\prod_{n=1}^N \sum_{z_n} p(z_n | \theta) p(w_n | z_n, \beta) \right) d\theta \quad (2-9)$$

$$p(\theta, Z | D, \alpha, \beta) = \frac{p(\theta, Z, D | \alpha, \beta)}{p(D | \alpha, \beta)} \quad (2-10)$$

但是，公式 (2-9) 的似然概率和公式 (2-10) 的后验概率的表达式中，积分中存在加和，无法计算^[32]，故采用变分推断。该方法的主要思想就是用一个简化的近似模型代替原始模型，通过让近似模型无限逼近原始模型，近似模型的参数值即可代替原始模型的参数。图 2-5 (b) 是近似模型的图模型（原始模型用 p 表示，近似模型用 q 表示）， γ 和 ϕ 是近似模型的参数，在给定观测变量 w 的条件下，假设 θ 和 z 独立，故图 2-5 (b) 中丢掉了节点 w 。近似模型 q 中的参数求解通过

最小化两个模型之间的距离，见公式 (2-11)，其中 \mathcal{D}_{KL} 表示近似模型 q 和原始模型 p 的 KL 散度（见公式 (2-12)），KL 散度是用来衡量两个分布之间的距离。在原始模型的似然函数表达式中引入近似模型的后验函数得到公式 (2-13)，其中，公式 (2-13) 的第二步到第三步不等式的转换过程中引入了詹森不等式，将公式 (2-13) 的右端定用 $\mathcal{L}(\gamma, \phi; \alpha, \beta)$ 表示称为近似的似然函数。把近似似然函数 $\mathcal{L}(\gamma, \phi; \alpha, \beta)$ 与两个模型的 KL 散度进行相加，得出近似似然函数和真实似然函数的差值正好等于两个模型的 KL 散度（见公式 (2-15)），过程见公式 (2-14)。

$$(\gamma, \phi) = \arg \min_{(\gamma, \phi)} \mathcal{D}_{KL}(q(\theta, z | \gamma, \phi) \| p(\theta, Z | D, \alpha, \beta)) \quad (2-11)$$

$$\begin{aligned} \mathcal{D}_{KL} &= \int \sum_Z q(\theta, Z | \gamma, \phi) \log \frac{q(\theta, Z | \gamma, \phi)}{p(\theta, Z | D, \alpha, \beta)} d\theta \\ &= \int \sum_Z q(\theta, Z | \gamma, \phi) \log q(\theta, Z | \gamma, \phi) d\theta \\ &\quad - \int \sum_Z q(\theta, Z | \gamma, \phi) \log p(\theta, Z | D, \alpha, \beta) d\theta \end{aligned} \quad (2-12)$$

$$\begin{aligned} \log p(D | \alpha, \beta) &= \log \int \sum_Z p(\theta, Z, D | \alpha, \beta) d\theta \\ &= \log \int \sum_Z q(\theta, Z | \gamma, \phi) \frac{p(\theta, Z, D | \alpha, \beta)}{q(\theta, Z | \gamma, \phi)} d\theta \\ &\geq \int \sum_Z q(\theta, Z | \gamma, \phi) \log \frac{p(\theta, Z, D | \alpha, \beta)}{q(\theta, Z | \gamma, \phi)} d\theta \\ &= \int \sum_Z q(\theta, Z | \gamma, \phi) \log p(\theta, Z, D | \alpha, \beta) d\theta \\ &\quad - \int \sum_Z q(\theta, Z | \gamma, \phi) \log q(\theta, Z | \gamma, \phi) d\theta \\ &= \mathcal{L}(\gamma, \phi; \alpha, \beta) \end{aligned} \quad (2-13)$$

在已知文档的情况下，公式 (2-15) 左边是固定值，最小化两个模型的 KL 散度相当于最大化近似似然函数 $\mathcal{L}(\gamma, \phi; \alpha, \beta)$ 。因此，通过以上的公式推导和分析，模型的求解最终变成最大化函数 $\mathcal{L}(\gamma, \phi; \alpha, \beta)$ 。根据模型变量之间条件概率的关系，目标函数 $\mathcal{L}(\gamma, \phi; \alpha, \beta)$ 简化为公式 (2-16)，对其进行极大值求解能够得到“主题-文档”分布 θ 和“主题-单词”分布 β 。

$$\begin{aligned}
\mathcal{L}(\gamma, \phi; \alpha, \beta) + \mathcal{D}_{KL} &= \int \sum_Z q(\theta, Z | \gamma, \phi) \log p(\theta, Z, D | \alpha, \beta) d\theta \\
&\quad - \int \sum_Z q(\theta, Z | \gamma, \phi) \log q(\theta, Z | \gamma, \phi) d\theta \\
&\quad + \int \sum_Z q(\theta, Z | \gamma, \phi) \log q(\theta, Z | \gamma, \phi) d\theta \\
&\quad - \int \sum_Z q(\theta, Z | \gamma, \phi) \log p(\theta, Z | D, \alpha, \beta) d\theta \\
&= \int \sum_Z q(\theta, Z | \gamma, \phi) \log p(\theta, Z, D | \alpha, \beta) d\theta \\
&\quad - \int \sum_Z q(\theta, Z | \gamma, \phi) \log p(\theta, Z | D, \alpha, \beta) d\theta \\
&= \int \sum_Z q(\theta, Z | \gamma, \phi) \log \frac{p(\theta, Z, D | \alpha, \beta)}{p(\theta, Z | D, \alpha, \beta)} d\theta \\
&= \int \sum_Z q(\theta, Z | \gamma, \phi) \log p(D | \alpha, \beta) d\theta \\
&= \log p(D | \alpha, \beta) \int \sum_Z q(\theta, Z | \gamma, \phi) d\theta \\
&= \log p(D | \alpha, \beta) \tag{2-14}
\end{aligned}$$

$$\log p(D | \alpha, \beta) = \mathcal{L}(\gamma, \phi; \alpha, \beta) + \mathcal{D}_{KL}(q(\theta, Z | \gamma, \phi) \| p(\theta, Z | D, \alpha, \beta)) \tag{2-15}$$

$$\begin{aligned}
\mathcal{L}(\gamma, \phi; \alpha, \beta) &= \int \sum_Z q(\theta, Z | \gamma, \phi) \log p(\theta, Z, D | \alpha, \beta) d\theta \\
&\quad - \int \sum_Z q(\theta, Z | \gamma, \phi) \log q(\theta, Z | \gamma, \phi) d\theta \\
&= E_q[\log p(\theta, Z, D | \alpha, \beta)] - E_q[\log q(\theta, Z | \gamma, \phi)] \\
&= E_q[\log(p(\theta | \alpha) p(Z | \theta) p(D | Z, \beta))] \\
&\quad - E_q[\log(q(\theta | \gamma) q(Z | \phi))] \tag{2-16} \\
&= E_q[\log p(\theta | \alpha) + \log p(Z | \theta) + \log p(D | Z, \beta)] \\
&\quad - E_q[\log(q(\theta | \gamma) + \log q(Z | \phi))] \\
&= E_q[\log p(\theta | \alpha)] + E_q[\log p(Z | \theta)] + E_q[\log p(D | Z, \beta)] \\
&\quad - E_q[\log(q(\theta | \gamma))] - E_q[\log q(Z | \phi)]
\end{aligned}$$

2.4.2 CTM

CTM^[35]是 LDA 模型的扩展。LDA 中假设每个文档的主题分布服从狄利克雷分布，导致了构成文档的主题之间相互独立，在实际的生活中，并不符合，例

如，一篇关于医学的文章，与遗传主题的关系一定比科技主题的关联更多。

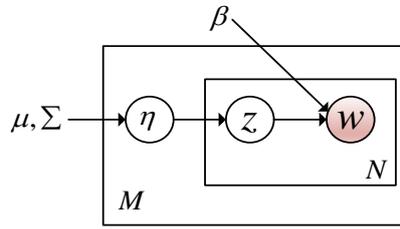


图 2-6 CTM 的图模型

CTM 中用高斯分布代替狄利克雷分布，图 2-6 是 CTM 的图模型，其中 μ 和 Σ 分别是高斯分布的均值和协方差，协方差能够刻画不同主题之间的关联，图模型中其他的变量和参数均和隐狄利克雷中的参数和变量相同。存在一个问题是，主题模型是对文档中离散的单词进行建模，狄利克雷分布是离散概率分布的，但是高斯分布是连续概率分布。CTM 中一个非常关键的技术是将连续分布函数离散化，离散化函数见公式 (2-17)，其中黑体的 η 算子表示通过高斯分布抽取的连续的文档主题分布，公式 (2-17) 将其进行了离散化和归一化。

$$\theta = f(\boldsymbol{\eta}) = \frac{\exp\{\boldsymbol{\eta}\}}{\sum_i \exp\{\eta_i\}} \quad (2-17)$$

CTM 中的参数是 μ 、 Σ 和 β ，求解方法和过程和 LDA 的方法类似，本小节不再进行重复介绍。

2.5 本章小结

本章节对于集群场景运动轨迹获取、片段轨迹预处理、轨迹建模的通用技术和模型进行了简单的介绍。第一节介绍了基于光流法的轨迹跟踪算法的基本原理。第二节叙述了集群场景中对于片段轨迹的预处理方法；第三节介绍了 BOW 字典构建的方法，其中包括对文本、图像以及轨迹三种数据的 BOW 字典的构建方法；最后一节介绍了自然语言分析领域中典型的主题模型，其中主要包括 LDA 和 CTM。通过对以上相关技术的介绍，为后续关于集群场景运动目标轨迹分类的研究进行了铺垫。

第三章 基于 BCTM 与多类 SVM 的集群目标轨迹分类

上一章介绍了与本文研究工作相关的技术，本章针对集群场景中运动目标轨迹分类问题进行研究，从运动目标轨迹的获取，轨迹预处理，到基于语义的中层特征学习和表达，最后对轨迹进行分类一整套流程进行详细的论述。

3.1 研究工作框架

图3-1是本章研究框图，阴影部分是创新点，具体研究内容叙述如下：

- 1) 首先采用KLT^[4]获取集群场景中运动轨迹；
- 2) 由于步骤1)中获取到的轨迹高度片断和混乱,无法直接进行分类分析,在MRF模型的基础上,提出Global-MRF有效捕捉原始的轨迹数据之间时空关联,这种时空上的关联能够表达集群场景中基于语义的分类特性；
- 3) 为了保证框架的可扩展性,本章采用BCTM对步骤2)中预处理后的轨迹进行建模,挖掘数据背后隐藏的基于语义的中层特征,重要的是这种基于语义的中层特征能够提供一种更加准确的表达方式；
- 4) 基于轨迹中层特征的学习,设计合理的特征表达方式和有效的分类流程,采用多类SVM (Multiclass Support Vector Machine, 多类支持向量机)^[36]分类器获得未知轨迹的类别标号。

图3-1是本章研究工作框架图，阴影部分是本章创新点。

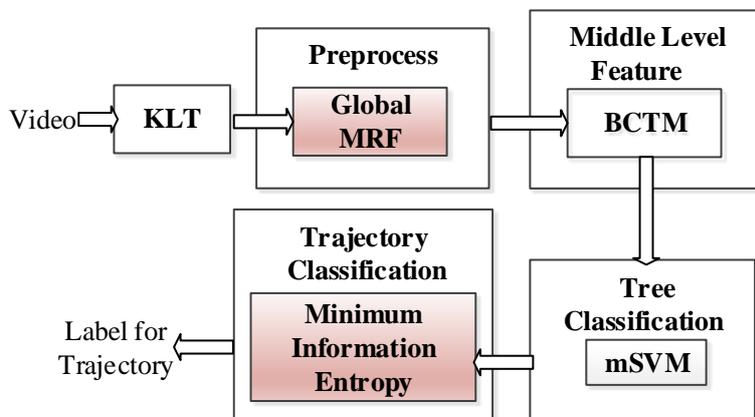


图 3-1 本章研究工作框架图

3.2 场景结构

正如本文的第一章所述，集群场景可以分为有结构的机群场景和无结构的集群场景。在有结构的集群场景中，具有固定的交通设施或者交通规则，能够约束场景中个体的运动行为，具有固定或者单一的运动模式，如马拉松、交通等集群场景；而在无结构的集群场景中，没有固定的交通设施或者交通规则约束场景中个体的运动行为，如中心广场等。但是，不论是有结构的集群场景还是无结构的集群场景都存在一些兴趣点，这些兴趣点大致可以分为两类，一类是显性兴趣点，如百货大厅进出口、售票点、报刊厅等，这些区域是场景中运动个体聚集、停留或者离开的区域；另一类是隐性兴趣点，如监控设备能够拍摄到的极点区域，这些区域并不是场景中固有的结构，而是因为监控设备的布局而出现的隐性结构信息，这些区域是场景中个体出现或者消失在监控视频中的区域。

不同的场景结构和监控设备的布局会形成不同的兴趣点，这些兴趣点能够协助分析场景中的运动模式，文献[9]中，定义运动轨迹出现和消失的区域定义为场景的进出口，采用混合高斯模型对其进行建模，学习其兴趣点区域。文献[4]中，把场景的进出口信息作为后续对轨迹进行建模的先验信息，有效提高场景中运动模式学习的准确性。文献[25]在文献[4]工作的基础上，提出 BCTM，模型中，进出口信息被假设服从多想是分布，轨迹的不完整造成不是所有轨迹都具有进出口信息，模型利用轨迹中已知的进出口信息，推断片段轨迹未知的进出口。

本文选择了两个不同类型的公开数据集，一个是纽约火车站场景^[5]，该场景属于无结构集群场景，另一个是交通场景^[6]，该场景存在固定的交通规则，是有结构的集群场景。在无结构的纽约火车站场景中，把明显的进出口区域作为该场景的兴趣点；在有结构的交通场景中，没有进出口区域，将视频拍摄到的极点位置定义为该场景的进出口，作为隐性的兴趣点。与文献[5]和文献[25]相似，本文根据经验定义两个场景中进出口区域，图 3-2 (a) 展示了纽约火车站的进出口区域，图中一共有 8 个椭圆形的进出口区域，图 3-2 (b) 展示了交通场景的进出口区域，图中一共有 5 个椭圆形的进出口区域图。

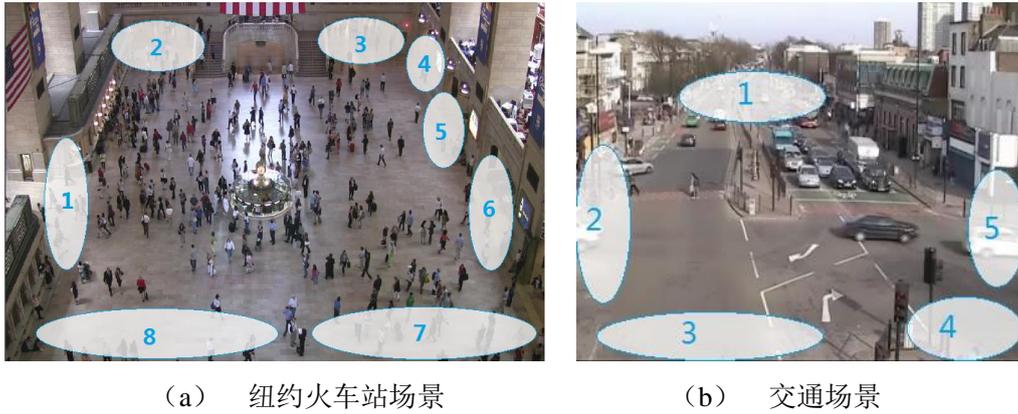


图 3-2 场景结构图

图 3-3 是一个进出口区域的定义示意图，图中以图像的中心为坐标中心，左下角椭圆表示一个进出口区域 $o^i(m, n)$ 是椭圆的中心坐标， a 、 b 分别为是椭圆的长短半轴（通常作为经验参数），进出口区域内的像素点必须满足公式 (3-1)，公式中下标 i 表示场景中第 i 个进出口。

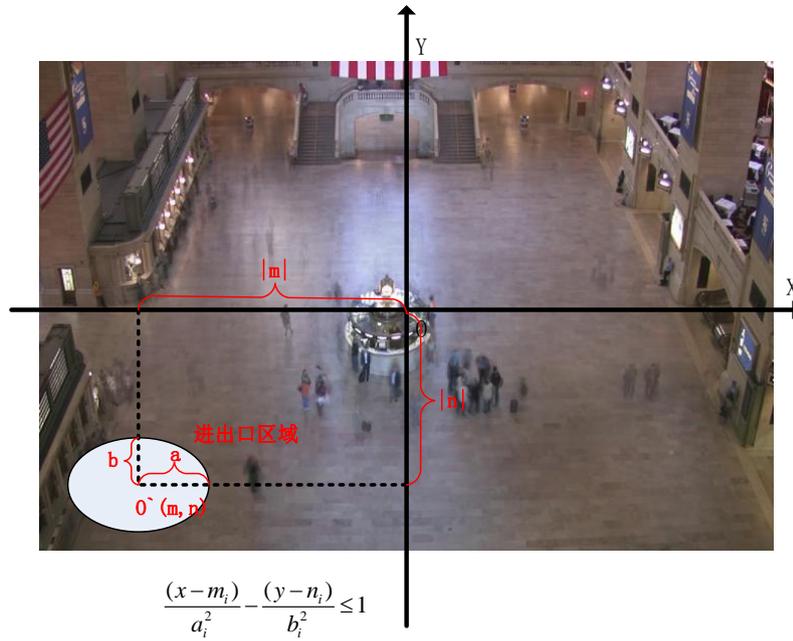


图 3-3 场景进出口区域划分示意图

$$\frac{(x-m_i)}{a_i^2} - \frac{(y-n_i)}{b_i^2} \leq 1 \quad (3-1)$$

3.3 运动目标轨迹提取

运动目标轨迹提取文研究工作数据获取的过程,本文选用 KLT 跟踪算法[6]。流向量序列是获取到的运动轨迹的表达方式,公式(3-2)中,轨迹 T 由连续的 N 个轨迹点构成,轨迹点 (x, y, t) 表示 t 时刻该运动目标的位置。

$$T = \{(x_1, y_1, t_1), (x_2, y_2, t_2), \dots, (x_N, y_N, t_N)\} \quad (3-2)$$

根据上个小节对场景结构的定义,轨迹可以被划分为完整轨迹、只有进口轨迹、只有出口轨迹和既没有进口也没有出口轨迹:

- 1) 完整轨迹: 轨迹具有完整的进出口信息, 即如公式 (3-2) 的轨迹表示中第一个轨迹点和最后一个轨迹点满足公式 (3-1), 如图 3-4 中红色线段所示;
- 2) 只有进口轨迹: 公式 (3-2) 的轨迹表达中只有起始轨迹点满足公式 (3-1), 无出口信息, 如图 3-4 中黄色线段所示;
- 3) 只有出口轨迹: 公式 (3-2) 的轨迹表达中只有终止轨迹点满足公式 (3-1), 无出口信息 (如图 3-4 中绿线段所示);
- 4) 既没有进口也没用出口轨迹: 公式 (3-2) 的轨迹表达中始发点和结束点都不满足公式 (3-1), 如图 3-4 中蓝色线段。

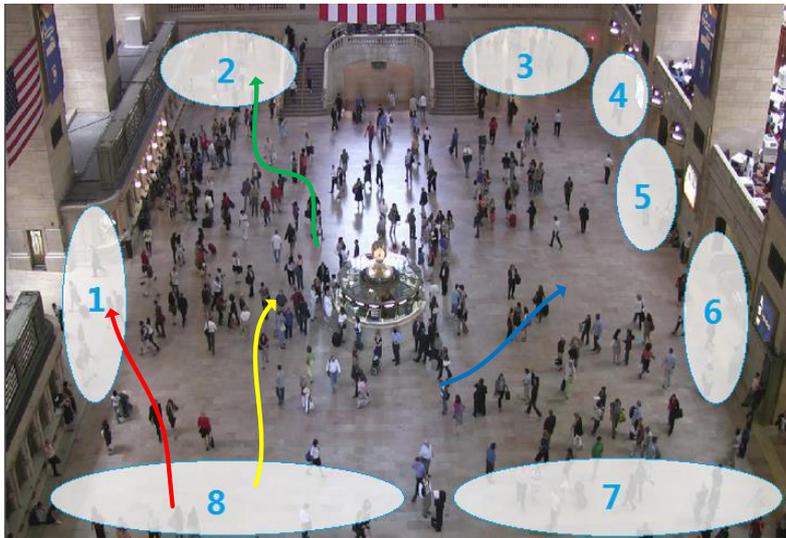


图 3-4 轨迹类别示意图

在公式 (3-2) 的基础上, 加入进出口信息, 每条轨迹被表达为公式 (3-3),

其中 s 、 h 分别是轨迹 T 的进出口的标号,如火车站场景为 1~8,交通场景为 1~5,若不存在进出口信息则用 0 表示。

$$T = \{(s, h), (x_1, y_1, t_1), (x_2, y_2, t_2), \dots, (x_N, y_N, t_N)\} \quad (3-3)$$

3.4 轨迹预处理

片段轨迹连接是轨迹的预处理过程。如本文第一章所述,集群场景中人流密度较大,运动个体之间遮挡严重,跟踪识别算法捕获到的运动轨迹片段性高,在轨迹的形状或者方向上不具有类别信息。但是,在集群场景中,不完整轨迹之间具有一定的时间和空间上的关联。文献[5]提出采用马尔科夫随机场模型进行建模,该模型用于挖掘轨迹之间的时空关联。本文在文献[5]的基础上提出了 Global MRF 进行处理。

3.4.1 Global MRF

MRF 模型的主要思想是把满足公式 (2-3) 定义的时空一致的邻域轨迹首尾进行连接,在连接的过程中,模型采用生成树的算法进行连接轨迹的搜索。轨迹首尾点的方向只是一个局部的运动方向,完全不能代表整条轨迹的运动趋势,另外,基于光流的跟踪算法对光照条件非常敏感,跟踪结果存在一定的噪音。因此,连接过程中仅仅满足当前连接轨迹首尾处的时空一致性,只有一个局部优化的作用,并不能保证全局的时空一致性。通过实验,发现 MRF 在连接轨迹的过程中会出现图 3-5 中黄色的绕圈连接,这种绕圈连接是一种错误连接,而且,这样的情况在运动复杂程度越高的场景中越明显。



图 3-5 MRF 中轨迹绕圈图

针对上述出现的问题，本文提出 Global MRF，主要的思想是根据连接轨迹的 K 邻域轨整体的平均速度方向迹控制和调整轨迹连接的全局一致性。具体的讲，轨迹 q 的初始轨迹点与轨迹 p 的终止轨迹点连接的时候，不仅要满足轨迹 q 起始点的速度方向和轨迹 p 终止点的速度方向一致，而且要满足与轨迹 q 的 k 个邻域轨迹的平均速度方向一致。如图 3-6 的示意图中，绿色代表轨迹 p ，黄色代表轨迹 q ，蓝色是轨迹 p 的邻域轨迹，未加入全局一致约束时候，绿色轨迹与红色轨迹起始点的速度方向比黄色轨迹起始点的速度方向更加一致，因此，会沿着红色轨迹片段连接成的完整轨迹；然而，加入了全局约束项后，红色轨迹和其邻域蓝色轨迹簇起始点的平均速度方向并不一致，通过这样的全局约束，绿色轨迹会沿着黄色路径进行连接。

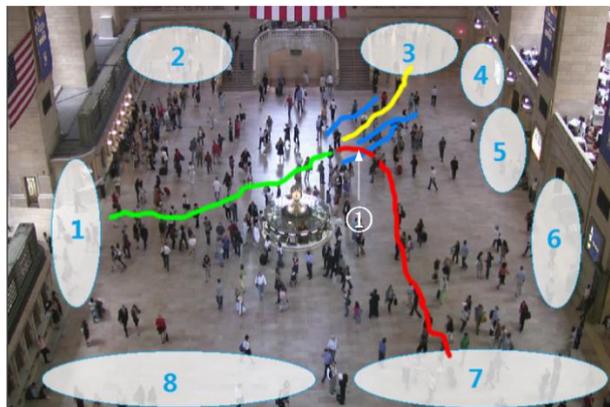


图 3-6 Global MRF 轨迹预处理示意图

基于文献[5]对 MRF 的定义，Global MRF 模型在能量函数的计算公式中加入了全局一致性约束项，具体的数学表达式见公式 (3-4)。

$$W(q, p) = \begin{cases} \exp\left[\left(\frac{v_q^s \cdot v_p^e}{\|v_p^e\| \cdot \|v_q^s\|} - 1\right) + \lambda \left(\frac{v_q^{avg} \cdot v_{knn(p)}^{avg}}{\|v_q^{avg}\| \cdot \|v_{knn(p)}^{avg}\|} - 1\right)\right], & \text{if (2-3) satisfies,} \\ 0, & \text{otherwise.} \end{cases} \quad (3-4)$$

其中，下标 $knn(p)$ 表示轨迹 p 的 k 个邻域轨迹，上标 avg 表示平均速度， $\lambda (\lambda > 0)$ 是控制全局一致性的参数，全局一致性程度随着 λ 的增大而增大。公式

(3-4) 中，指数函数中前一项中 $\frac{v_q^s \cdot v_p^e}{\|v_q^s\| \cdot \|v_p^e\|}$ 保证了轨迹 q 、 p 首尾的速度方向一

致，后一项中 $\frac{v_q^{avg} \cdot v_{knn(p)}^{avg}}{\|v_q^{avg}\| \cdot \|v_{knn(p)}^{avg}\|}$ 保证了轨迹 q 在全局上的一致性。

3.4.2 实验结果与分析

本节给出了 Global MRF 和 MRF 对原始轨迹预处理的实验结果。实验数据来自 KLT 从纽约火车站场景监控视频中捕获的运动轨迹集，其中监控视频来自文献[5]。图 3-7 展示了 Global MRF 和 MRF 轨迹连接的实验结果图，其中 (a) (c) (e) (g) 为 Global MRF 的结果，(b) (d) (f) (h) 是 MRF 的结果。



(a) Global MRF



(b) MRF

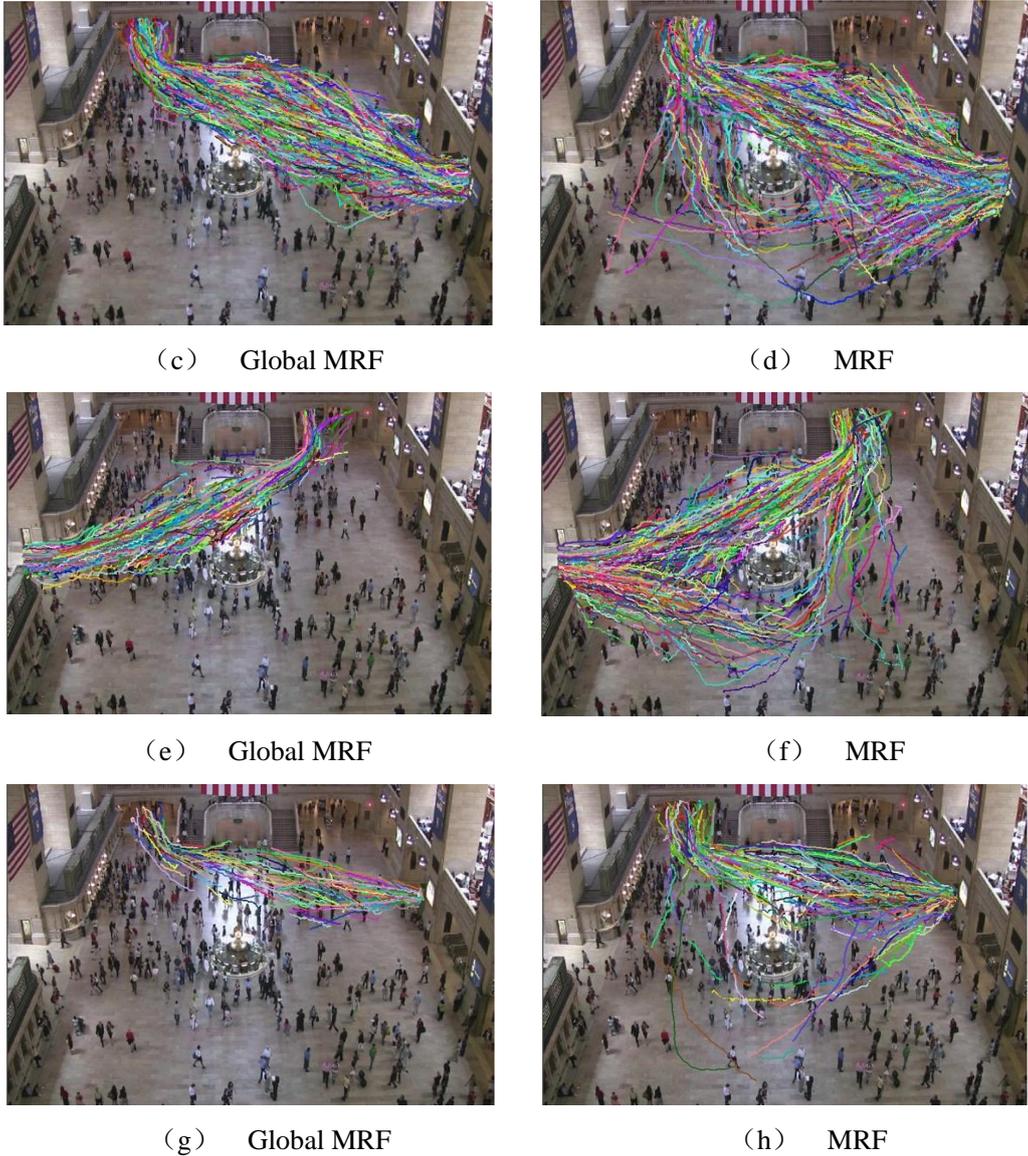


图 3-7 Global MRF 和 MRF 实验结果

3.5 中层特征学习

预处理后，片段、混杂的原始轨迹依照时空关联被连接成完整的轨迹树，这样的轨迹树虽然能够从形状或者方向上具备一定的类别特性，但是，仍然不能直接用来分类，具体原因如下：

- 1) 轨迹维度较高，不能直接用 SVM (Support Vector Machine, 支持向量机)^[37]等分类器直接进行分类；轨迹底层特征字典的维度和视频的分辨率相同，

以纽约火车站场景为例，图像分辨率为 720×480 ，则轨迹底层特征字典维度为 345600，每条轨迹在字典中的特征表达非常稀疏，统计可得，每个轨迹样本在 345600 特征维度上仅仅只有不到 300 个特征元素上有数据，这种高度稀疏的数据并不适用于直接进行分类；

2) 轨迹预处理是依据片断轨迹之间时空关联决定是否进行轨迹连接，在这个过程中并不关心被连接的轨迹是否是同一个运动目标的运动轨迹，并且，一条片段的轨迹很可能被多条轨迹树所共享，对完整轨迹树直接进行分类还是不能达到对单条轨迹进行分类的目的。如图 3-8，红色片段轨迹被分别连接到黄色轨迹树和绿色轨迹树中。

基于现有的集群场景的理论，集群场景中运动个体的运动行为逐渐被“从众”的运动模式所替代^[5,29]。这种运动模式可以看作隐性的特征，文献[5]中把其称作基于语义的运动区域。在无结构的运动场景中，“从众”的运动模式会形成隐形的通道。这种基于语义的特征能够更准确的对轨迹进行表达和描述，这一点对于分类准确程度至关重要。另外，用基于语义的中层特征对轨迹进行重描述保证了框架不依赖底层数据，使框架具有可扩展性。在本章的研究工作中，选择了 BCTM^[25]，下面将对 BCTM 进行详细介绍。

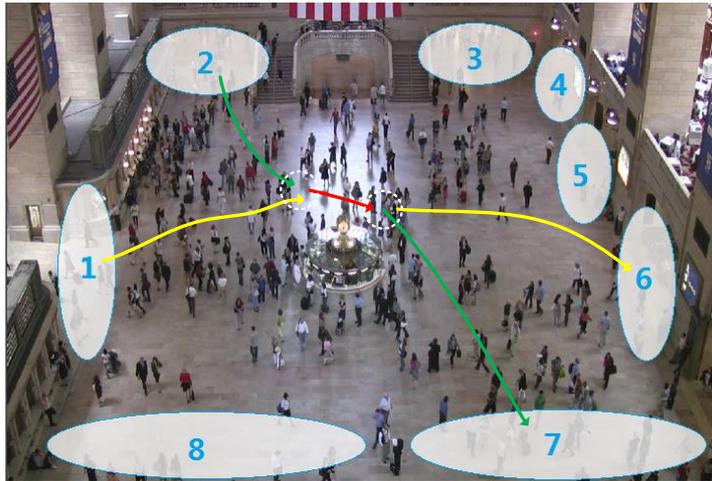


图 3-8 片断轨迹被连接到不同的轨迹树中

3.5.1 BCTM

BCTM^[25,26]是由邹等研究者提出,用于挖掘集群场景中运动轨迹背后隐藏的基于语义的中层特征。该模型在CTM的基础上加入了置信算子,让模型挖掘到的中层特征更具有更强的判别力,这里的置信是指运动轨迹在集群场景中的进出口信息。

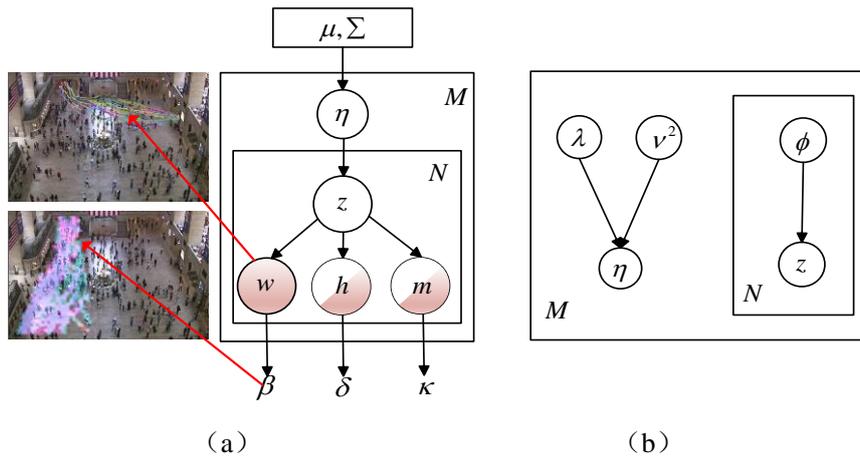


图 3-9 BCTM 的图模型 (a) 和近似模型 (b)

一般的将主题模型用于解决计算机视觉问题都要具备概念上的对应, BCTM 作为主题模型的扩展模型也不例外。该模型用于解决集群运动轨迹分析中, 轨迹点 w 被视为主题模型中的单词, 表示轨迹相当于文档, 用 $D = (w_1, w_2, \dots, w_N)$ 表示, 全部的轨迹数据相当于语料库, 表示为 $Corpus = (D_1, D_2, \dots, D_M)$, 数据背后隐藏的基于语义的中层特征等价于主题模型中的主题, 这种基于语义的中层特征可视化在场景中如图 3-9 中, 集群场景中运动个体从左下角到右上角经过的通道。

图 3-9 是 BCTM 的图模型, 图模型中参数和变量说明如下:

- 1) μ 和 Σ 分别是高斯分布的均值和协方差;
- 2) 主题分布变量 η 服从高斯分布, 即 $\eta \sim N(\mu, \Sigma)$;
- 3) 主题变量是隐变量服从参数为 θ 的多项式分布, 即 $z \sim Multinomial(\theta)$,

$$\text{其中 } \theta = f(\eta) = \frac{\exp\{\eta\}}{\sum_i \exp\{\eta_i\}};$$

4) 可观测变量 w 表示词项, 服从 z 条件下参数为 β 的多项式分布, 即 $p(w|z_w, \beta) \sim \text{Multinomial}(\beta_{z_w})$; 部分阴影变量 h 和 m 表示半观测变量, 也就是说有可能观测到也可能观测不到, 这两个变量表示置信信息, 分别代表轨迹的进口和出口, 分别服从 z 条件下参数为 δ 、 κ 的多项式分布, 即 $p(h|z_h, \delta) \sim \text{Multinomial}(\delta_{z_h})$ 、 $p(m|z_m, \kappa) \sim \text{Multinomial}(\kappa_{z_m})$;

5) 超参数 β 是 $K * V$ 的矩阵, K 是主题的个数, V 是“词包”字典中单词的个数, β_{ij} 表示字典中第 j 个单词属于主题 i 的概率, β_i 表示第 i 个主题的单词分布; 超参数 δ 是置信主题概率算子, 与词项类似, 每个置信信息都有其主题分布, 故 δ 定义为 $K * H$ 的矩阵, K 是主题的个数, H 是场景中进口区域的个数, δ_{ij} 表示进口 j 属于主题 i 的概率, δ_i 表示主题 i 的进口 $K * S$ 分布; 同样的, 超参数 κ 是 $K * S$ 的矩阵, K 是主题的个数, S 是场景中进口区域的个数, κ_{ij} 表示出口 j 属于主题 i 的概率, κ_i 表示主题 i 的出口分布。

3.5.2 模型公式推导

3.5.1 节介绍了 BCTM 图模型及其定义, 在此基础上, 对模型中参数推导进行介绍。

对于一条轨迹, 随机变量 η 、 Z 、 D 、 H 和 M 的联合概率为公式 (3-5), 其中 z_n 表示轨迹 D 中第 n 个轨迹点的主题, 将公式 (3-5) 中的变量 η 和 Z 积分加和后, 得到模型的似然函数表达式 (3-6), 公式 (3-7) 是模型后验概率表达式。

$$p(D, H, M, \eta, Z | \mu, \Sigma, \beta, \delta, \kappa) = p(\eta | \mu, \Sigma) \prod_{n=1}^N p(z_n | f(\eta)) p(w_n | z_n, \beta) p(h | z_n, \delta) p(m | z_n, \kappa) \quad (3-5)$$

$$p(D, H, M | \mu, \Sigma, \beta, \delta, \kappa) = \int p(\eta | \mu, \Sigma) \prod_{n=1}^N \sum_{z_n} p(z_n | f(\eta)) p(w_n | z_n, \beta) p(h | z_n, \delta) p(m | z_n, \kappa) d\eta \quad (3-6)$$

$$p(\eta, Z | D, \mu, \Sigma, \beta, \delta, \kappa) = \frac{p(D, H, M, \eta, Z | \mu, \Sigma, \beta, \delta, \kappa)}{p(D, H, M | \mu, \Sigma, \beta, \delta, \kappa)} \quad (3-7)$$

模型的似然函数和后验概率难以计算，因此，模型采用变分推断进行模型参数推导，图 3-9 (b) 是变分推断的近似模型的图模型。模型公式推导的思想与 2.4.1 章节介绍的 LDA 的推导思想相同，真实模型的似然概率函数与近似的似然概率函数的差值与两个模型的 KL 散度相同，故模型的优化函数变为最大化函数 \mathcal{L} ，具体表达公式见 (3-8)

$$\begin{aligned} \mathcal{L} = & E_q[\log p(\eta | \mu, \Sigma)] + E_q[\log p(Z | \eta)] + E_q[\log p(D | Z, \beta)] \\ & + E_q[\log p(H | Z, \delta)] + E_q[\log p(M | Z, \kappa)] \\ & - E_q[\log q(\theta | \lambda, \nu^2)] - E_q[\log q(Z | \phi)] \end{aligned} \quad (3-8)$$

最大化目标函数 \mathcal{L} 解得模型变量和参数的表达式如下公式 (3-9)：

$$\begin{aligned} I. \quad & \phi_{ki} \propto \exp(\lambda_j) * \beta_{ki} * \delta_{ki} * \kappa_{ki} \\ II. \quad & \beta_k \propto \sum_i \phi_{ki} n_{w_i} \\ III. \quad & \delta_k \propto \sum_i \phi_{ki} n_{h_i} \\ IV. \quad & \kappa_k \propto \sum_i \phi_{ki} n_{m_i} \end{aligned} \quad (3-9)$$

其中，变量 ϕ_{ki} 表示轨迹点的中层特征分布，即轨迹 j 的第 i 个轨迹点属于中层特征 k 的概率是 ϕ_{ki} ；超参数 β_k 表示第 k 个中层特征， β_{ki} 表示轨迹点 i 属于中层特征 k 的概率， n_{w_i} 表示轨迹 j 中出现轨迹点 i 的个数；超参数 δ_k 表示中层特征 k 的进口置信分布， δ_{ki} 表示轨迹点 i 的进口置信属于中层特征 k 的概率， n_{h_i} 表示轨迹 j 中出现进口 h 的轨迹点的个数；超参数 κ_k 表示中层特征 k 的出口置信分布， κ_{ki} 表示轨迹点 i 的出口置信属于中层特征 k 的概率， n_{m_i} 表示轨迹 j 中出现进口 h 的轨迹点的个数。公式 (3-9) 中可以明显的看出，置信算子 δ 和 κ 加入了更新词项主题分布算子 ϕ 的更新过程中，通过 ϕ 算子影响特征算子 β 的学习过程。

3.5.3 轨迹重描述

BCTM 能够获取运动轨迹的中层特征空间和轨迹点的中层特征分布，分别

用算子 β 和 ϕ_i 表示。算子 ϕ_{ki} 表示轨迹点 i 属于主题 k 的概率，定义轨迹点 i 的中层特征表示为 $\arg \max_k \phi_{ki}$ 。包含 n 个轨迹点的轨迹 j 用 BCTM 学习到的中层特征进行重描述，具体数学表达式见公式 (3-10)，其中 f_i 表示轨迹点 i 的中层特征

$$\begin{aligned} T_j &= (f_1, f_2, \dots, f_n) \\ f_i &= \arg \max_k \phi_{ki} \end{aligned} \quad (3-10)$$

3.6 轨迹分类

基于轨迹的重描述，选用多类 SVM^[36]分类器，设计合理的分类流程对每条轨迹进行分类。

3.6.1 多类支持向量机

多类 SVM 是针对多分类问题提出的分类算法，该算法继承了 SVM^[37,38]找最大分类边界的特性，同时解决了“*One aginst All*”^[39]算法分类盲区的问题。

传统的 SVM 分类器是二值分类器，利用其解决多类数据分类问题的方法被称作“*One aginst All*”。该算法的思想是训练时依次把某个类别归为一类，其他剩余的样本归为另一类， k 个类别的样本训练出 k 个 SVM 分类器，针对第 k 类训练出来的分类器只能够对未知的样本进行属于第 k 类或者属于其他类进行判断，这样的分类算法存在一定的分类盲区。如图 3-10 中，红色圆圈、绿色矩形和蓝色三角形分别代表三类样本，虚线代表分类器，不同颜色是针对不同类别样本训练的分类器，比如红色虚线是针对红色圆圈样本训练的分类器，能够判断一个未知样本是否属于红色圆圈类别，若未知样本不属于红色圆圈类别，该分类器不能判断到底是属于蓝色三角类别还是属于绿色矩形类别。阴影部分区域是分类盲区，若有未知样本落入该区域，“*One aginst All*”将无法区分样本的类别。多类 SVM 分类器能够完全避免这样的分类盲区，图 3-10 中黑色实线表示多类 SVM 的分类面。

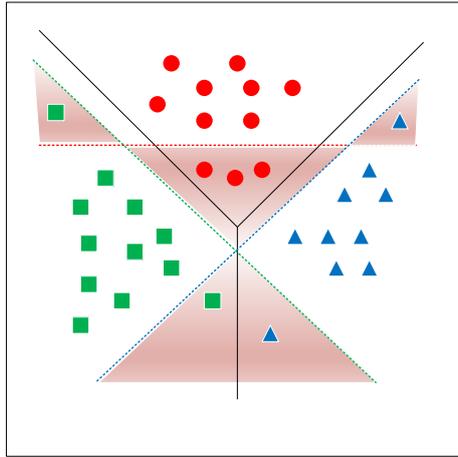


图 3-10 多类 SVM 与 “One against All” 对比图

3.6.2 分类流程的设计与实现

为了达到对每条轨迹分类的目的，设计如下的分类计算流程：

1) 构造分类器的输入：由于多类 SVM 输入的样本长度相同，故对每条轨迹的中层特征表达进行直方图统计。包含 m 个样本的样本集表示见公式(3-11)，公式中 x_i ($i \in [1, m]$) 代表样本， y_i ($i \in [1, m]$) 代表样本 i 的标号，样本 x_i 的数学表示见公式 (3-12)，其中 K 是 BCTM 学到的中层特征的个数， n_k 表示样本 i 中出现中层特征 k 的次数。

$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\} \quad (3-11)$$

$$x_i = (n_1, n_2, \dots, n_K) \quad (3-12)$$

2) 片段轨迹的分类：上述分类器能够得到未知的预处理后轨迹树的类别标号，但是并没有达到对每一条未知的片段轨迹分类的目的，这是因为正如 3.5 节中论述到一条片段轨迹可能被连接到不同的轨迹树中，那么一条片段轨迹很可能具有两个类别标号。为了得到对每条轨迹分类的目的，利用基于置信关联主题模型输出的 ϕ_i 算子，采用最小信息熵来为每条轨迹确定一个唯一的类别标号，类别标号公式见 (3-13)。

$$\begin{aligned}
 y_j &= \arg \min_{p \in tree(j)} H(p) \\
 H(p) &= -\sum_{i=1}^n \phi_i^{y_p} \log \phi_i^{y_p}
 \end{aligned} \tag{3-13}$$

其中, y_j 表示轨迹 j 类别标号, $tree(j)$ 表示轨迹 j 所属于的轨迹树集合, $H(p)$ 表示完整轨迹树 p 的信息熵, $\phi_i^{y_p}$ 表示轨迹树 p 中轨迹点 i 属于中层特征 y_p 的概率, y_p 表示轨迹树 p 的类别。

3.7 实验结果与分析

本节将展示上述基于多类 SVM 与 BCTM 的集群目标轨迹分类方法进行的实验结果与分析。本节选取了两种不同类型的公开数据集进行实验, 分别是纽约火车站场景^[5]和交通场景^[6], 纽约火车站场景以人群为运动目标的无结构室内场景, 交通场景是以车辆为运动目标的有结构的室外场景。具体的实验过程如下:

1) 采用 KLT 跟踪器分别从纽约火车站场景和交通场景的数据集中提取轨迹数据, 两个数据集的轨迹个数大约分别为 50000 条和 20000 条;

2) 用 Global MRF 对 1) 中提取的片段的运动轨迹进行预处理, 形成完整的运动轨迹树;

3) 分别对纽约火车站数据集和交通数据集进行 BOW 字典构建, 具体的, 纽约火车站场景数据集的视频像素为 720×480 , 构建的字典维度为 $(720/10) \times (480/10) \times 4 = 13824$; 交通场景数据集的视频像素为 290×360 , 构建的字典维度为 $(290/10) \times (360/10) \times 4 = 4176$;

4) 基于 3) 中 BOW 数据的构建, 采用 BCTM 学习运动目标轨迹的中层特征, 我们选择了 LDA 模型和 CTM 作为实验的对比模型。根据不同的数据集, 设定了不同个数的中层特征(主题), 具体的, 纽约火车站场景中中层特征(主题)的个数从 5 变化到 90, 交通场景中中层特征(主题)个数从 5 变化到 30;

5) 基于 4) 中对运动目标轨迹中层特征的学习, 用多类 SVM 进行轨迹分类, 我们对纽约火车站场景和交通场景数据集进行了人工数据标定, 分别标定了

3000 多条轨迹和 600 多条轨迹进行轨迹分类准确率的评测实验。

我们选用 LDA 和 CTM 作为 BCTM 的实验对比方法，分别选取了三个模型学习的部分中层特征进行可视化，图 3-11，3-12 分别是纽约火车站场景数据集和交通场景数据集的中层特征学习的实验对比，其中 (a) (b) 是 BCTM 的实验结果，(c) (d) 是 CTM 的实验结果，(e) (f) 是 LDA 的实验结果。

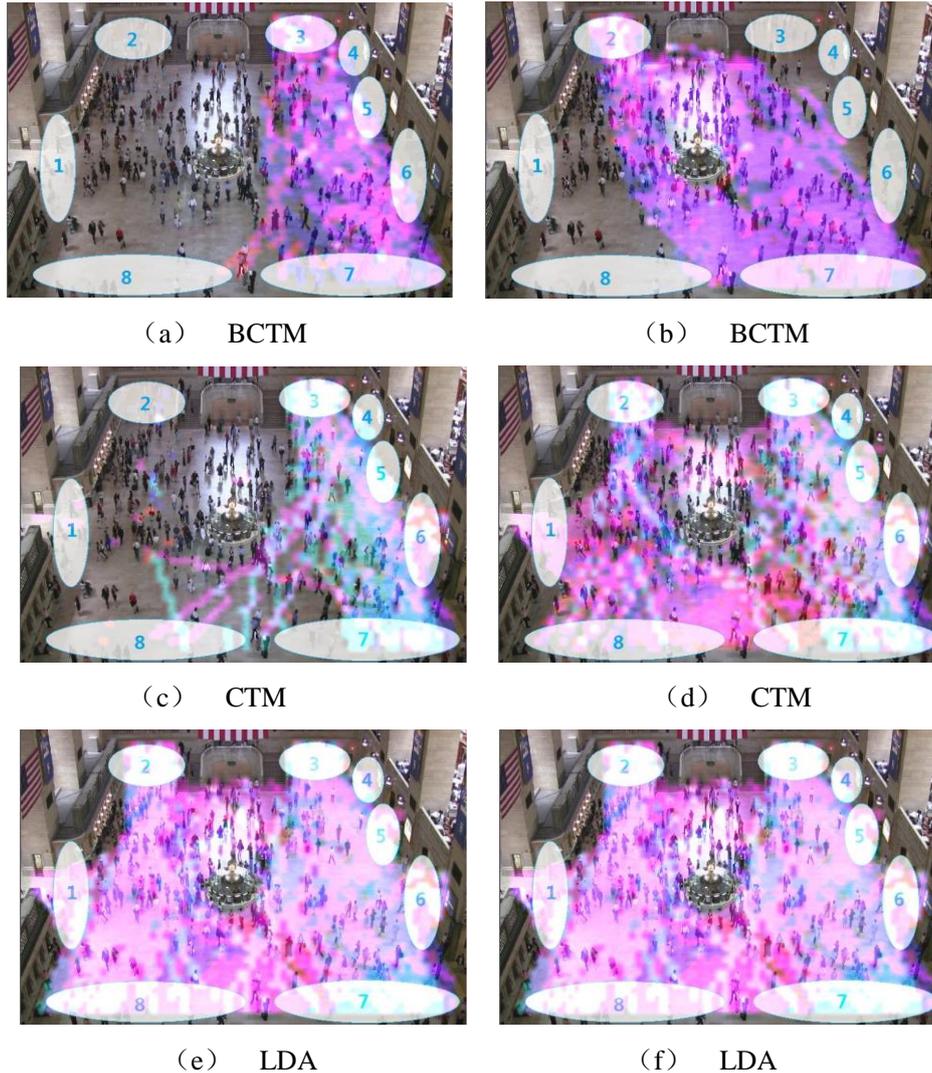


图 3-11 纽约火车站数据集中层特征可视化实验结果（不同的颜色表示不同中层特征的词项可视化，颜色的深浅程度代表该词项的个数）

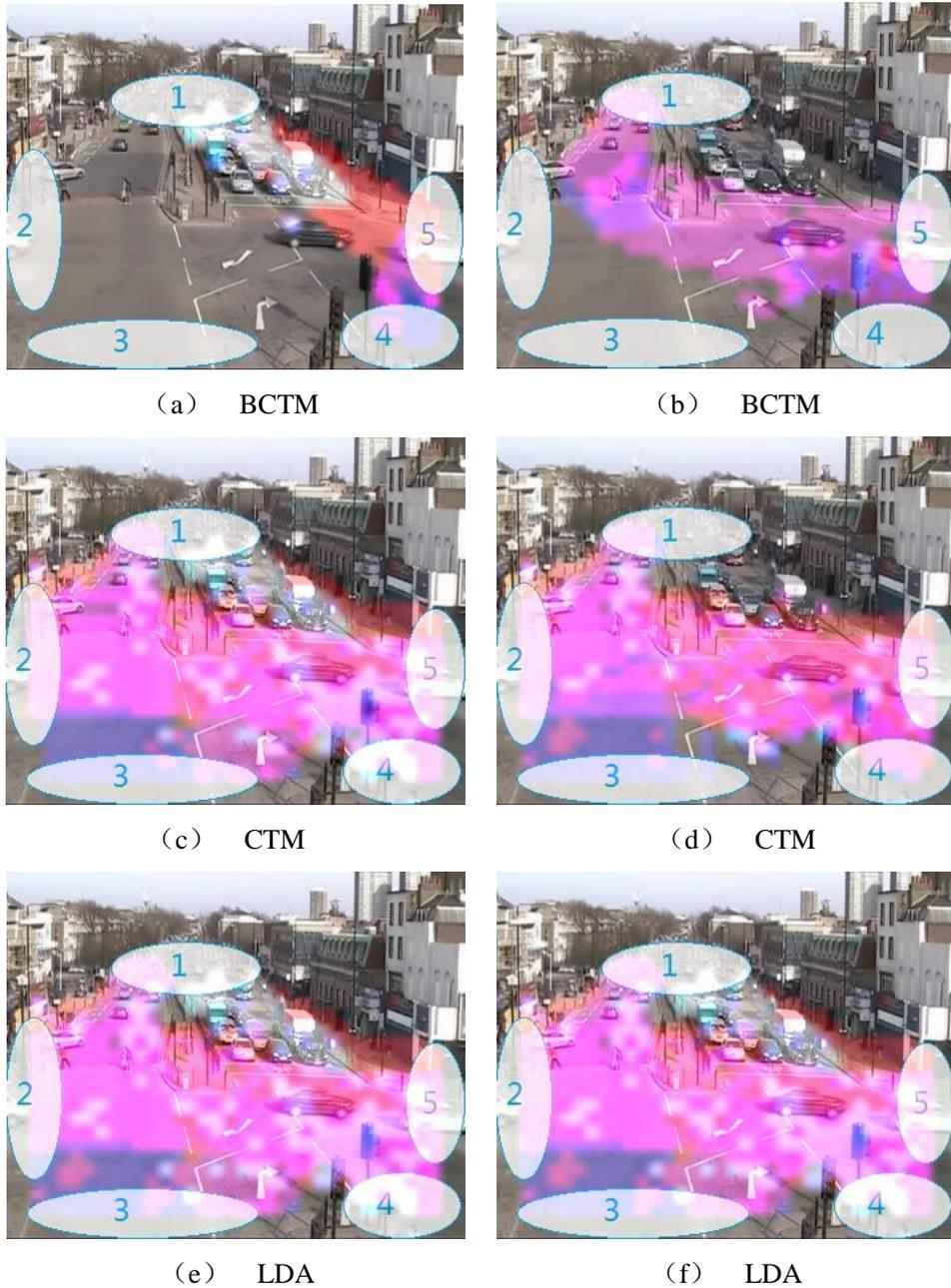


图 3-12 交通数据集中层特征可视化实验结果（不同的颜色表示不同中层特征的词项可视化，颜色的深浅程度代表该词项的个数）

从上述中层特征提取的实验结果中可以看出，BCTM 学习到的中层特征具有更强的判别力，下面我们对结果进行分析。

从模型的角度分析这一实验结果，CTM 与 LDA 模型的区别是：用高斯分布替换了 LDA 中的狄利克雷分布，更准确的刻画不同主题之间的关联，基于 BCTM

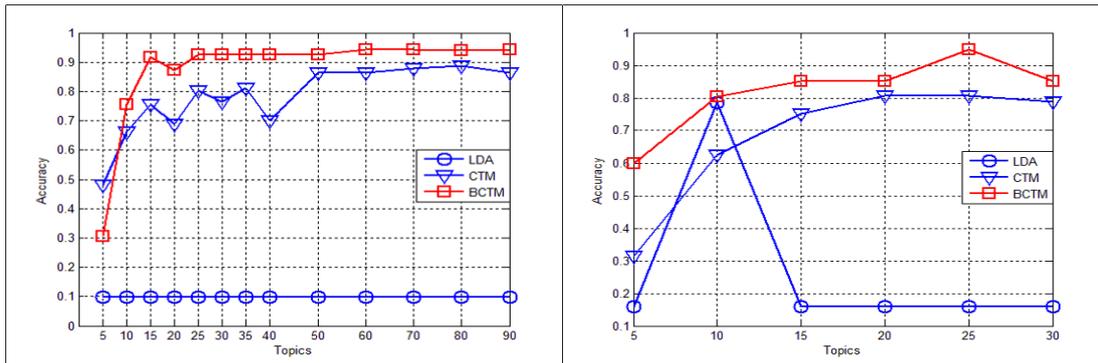
在 CTM 的基础上把运动目标轨迹的新出口置信信息作为先验加入到模型当中，增强了模型的学习能力。

从理论上分析，BCTM、CTM 和 LDA 模型的中层特征描述算子 β 的迭代更新公式相同，数学表达式见公式 (3-14)。但是，在 BCTM 中词项-主题概率算子 ϕ_{ki} 迭代更新公式是公式 (3-15) 的 I，而 CTM 和 LDA 模型的词项-主题概率算子 ϕ_{ki} 迭代更新公式是公式 (3-15) 的 II。I 与 II 的区别是：II 中加入了进口置信描述子 δ 和出口置信描述子 κ ，使得模型具有更强的学习、描述能力。

$$\beta_k \propto \sum_i \phi_{ki} n_{w_i} \quad (3-14)$$

$$\begin{aligned} I. \quad \phi_{ki} &\propto \exp(\lambda_j) * \beta_{ki} * \delta_{ki} * \kappa_{ki} \\ II. \quad \phi_{ki} &\propto \exp(\lambda_j) * \beta_{ki} \end{aligned} \quad (3-15)$$

图 3-13 是轨迹分类的准确率，其中 (a) 是纽约火车站场景数据集，(b) 是交通场景数据集。图中横坐标表示中层特征（主题）个数，纵坐标表示分类的准确率，红色带矩形线段表示用 BCTM 学习中层特征进行轨迹分类的准确率，蓝色带三角线段表示用 CTM 学习中层特征进行轨迹分类的准确率，蓝色带圆形线段表示用 LDA 模型学习中层特征进行轨迹分类的准确率。从图中可以看出，BCTM 学习的中层特征具有更准确的表达能力，分类的准确率更高且性能稳定。



(a) 纽约火车站场景

(b) 交通场景

图 3-13 分类准确率曲线图

3.8 本章小结

本章节对基于 BCTM 与多类 SVM 的集群目标轨迹分类方法的设计和实现进行了详细的叙述，主要对集群场景中轨迹获取、轨迹底层特征表达、轨迹预处理、中层特征学习与描述以及轨迹分类流程设计进行详细的介绍。通过以上各个环节的设计与分析，实现了对集群场景中片段轨迹进行分类的目的，同时，在两个公开的数据集上对本章节的方法进行了实验验证，具有一定的可行性和稳定性。

第四章 基于 Multi-CTM 集群目标轨迹分类

第三章详细论述了基于 BCTM 与多类 SVM 的集群目标轨迹分类框架的设计和实现,本章是上一章研究工作的并行工作,着重考虑轨迹分类问题在现实中人工标定费时费力的困难,借助生成扩展的模型改进思想,提出 Multi-CTM,设计并实现了弱监督的集群目标轨迹分类框架。

4.1 主题模型与分类

主题模型最早起源于自然语言处理领域用于挖掘文档主题的一种方法,近年来被广泛应用于计算机视觉领域。最基础 LDA 是一种非参数的贝叶斯模型 (Nonparametric Bayesian Model), 由于模型的非监督学习过程, LDA 不能直接用于解决分类问题。为了利用主题模型解决分类问题, 文本处理领域和计算机视觉领域的研究者提出一些 LDA 的扩展模型, 按照扩展的方法可将这些方法可以分为判别扩展 (Discriminant Extensions) 和生成扩展 (Generative Extensions)^[40,41]。

判别扩展的思想是在主题模型完成学习和表达后, 结合判别分类器进行分类, 如 SVM^[37-39], Multi-SVM^[36]。在文献^[42]中, 每张图像用 PLSA 模型学习的主题向量进行表达后用 SVM 分类器进行图像分类, 可以称为 SVM-PLSA^[40-42]。

生成扩展是指在原始的 LDA 模型基础上扩展为监督主题模型^[40-43]。Blei 等研究者在文献^[43]中首次提出 sLDA (Supervised Latent Dirichlet Allocation, 监督隐狄利克雷分布) 进行文档预测。Blei 等的研究工作之后, 相继提出了用监督主题模型解决解决计算机视觉问题。文献^[44]将 sLDA 模型应用于图像标注的工作中。Li-Feifei 等研究者在文献^[31]中提出 classLDA (Class-Latent Dirichlet Allocation, 基于类别的隐狄利克雷分布) 解决图像分类问题。

第三章的研究工作正是借助判别扩展的思想, 结合 BCTM 和多类 SVM 对未知轨迹进行分类。本章的研究工作则是借助生成扩展的思想, 提出 multi-CTM 对集群目标轨迹进行分类。这个模型的思想借鉴文献^[31]中 classLDA 的思想, 下一节将对 classLDA 模型的理论基础进行论述。

4.2 classLDA

ClassLDA^[31]是 Li-Feifei 等人在 2005 年提出，用于解决图像分类问题，该模型的根本思想是用多个 LDA 模型对图像数据进行基于语义的特征学习，每个 LDA 模型用不同类别的数据训练，若有 C 个类别的图像数据，模型训练的过程相当于分别训练了 C 个不同的 LDA 模型，每个 LDA 模型相当于一个分类器，对于未知类别的图像数据，认为与其最匹配的分类器对应的类别即为该未知图像类别。文献[31]的工作中，一副图像包含 n 个 BOW 特征，每个 BOW 特征被视为一个单词，图像被视为文档。图 4-1 是 classLDA 的图模型。与基础的 LDA 模型不同的是，classLDA 模型中有 C 组超参数 α 和 β ， c 表示类别标号，是可观测变量。模型中假设参数分布如下：

- 1) 主题分布变量 θ 服从类别参数 c 条件下参数为 α 的狄利克雷分布，即 $\theta \sim Dir(\alpha | c)$ ；
- 2) 主题变量 z 服从参数为 θ 的多项式分布，即 $z \sim Multinomial(\theta)$ ；
- 3) 超参数 β 是 C 个 $K * V$ 矩阵集合，每个矩阵对应一个主题特征算子；
- 4) 观测变量单词 w 服从主题变量 z_w 条件下参数为 β^c 的多项分布，即 $p(w | z_w, \beta^c) \sim Multinomial(\beta_{z_w}^c)$ ；

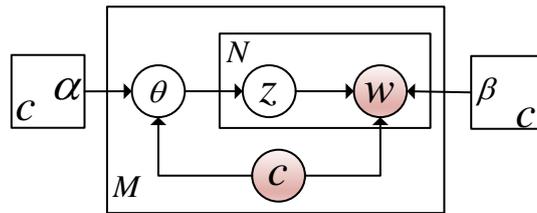


图 4-1 classLDA 的图模型

在模型数据分布假设的基础上，模型的似然概率函数 $p(x|\alpha, \beta, c)$ 的数学表达式见公式 (4-1)，模型的参数和变量的求解仍然采用变分推断的方法，具体的数学公式推导过程与 LDA 和 BCTM 相似，分别在 2.4.1 和 3.5.2 中详细论述。

$$p(D|\alpha, \beta, c) = \int p(\theta|\alpha, c) \prod_{n=1}^N \sum_{z_n} p(z_n|\theta) p(w_n|\beta_{z_n}^c) d\theta \quad (4-1)$$

对于一个未知类别的图像 D ，用已训练好的 C 组参数对图像 D 的似然概率函数进行求解，认为最大似然对应的类别即为图像 D 的类别，具体的数学公式为 (4-2)。

$$c = \arg \max_c p(D|\alpha, \beta, c) \quad (4-2)$$

classLDA 模型将不同类别的数据用不同的 LDA 模型学习，每个 LDA 模型学习出来的主题都是针对每一类数据的准确描述，能够有效的提高分类的准确率。文献[47]正是采用 classLDA 模型解决远距离的场景分类。但是，classLDA 模型解决图像分类问题，存在两个问题：

1) 模型的根本思想是将不同类型的数据用不同的 LDA 模型训练，这样的训练方式强制的切断了不同类别数据之间的关联，这样的假设实际上并不合理。图 4-2^[48] (a) (b) (c) 的类别分别是海岸、雪山和田野，虽然三副场景图片的类别不同，但是它们都包含了“天空”这个元素，并且占据了大部分的图像面积，针对这样的问题，强制的切断不同类别数据之间的关联并不能够带来有效的特征学习和表达效果；



(a) 海岸

(b) 雪山

(c) 田野

图 4-2 不同类别的场景包含相同元素“天空”

2) classLDA 模型实质上就是用多个 LDA 子模型训练不同类别的数据，子模型的个数和训练数据的类别数相同，这就造成 classLDA 模型不能解决类别数过大的数据集的分类问题。

4.3 Multi-CTM

针对集群轨迹分类，上述的两个问题并不存在。对于问题 1)，直观的讲，

章基于多类 SVM 与 BCTM 的集群目标轨迹分类框架中轨迹获取、轨迹预处理的基础上，设计并实现基于 Multi-CTM 集群目标轨迹分类，图 4-3 是基于 Multi-CTM 集群目标轨迹分类的框架流程图。

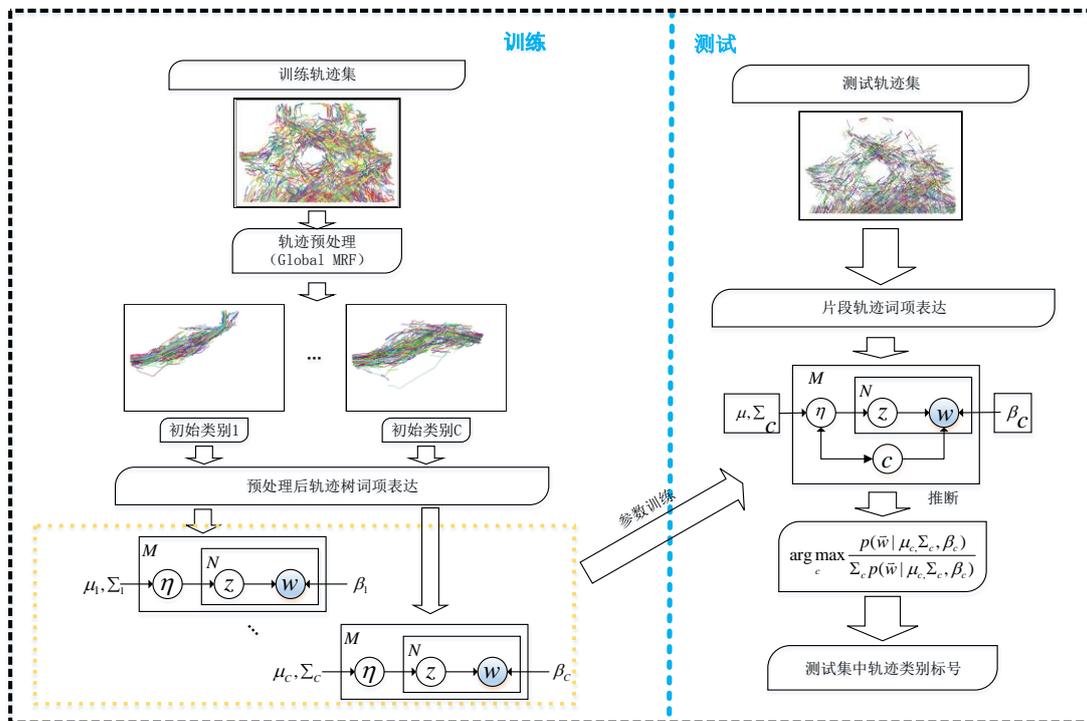


图 4-4 研究工作框架图

基于 Multi-CTM 集群目标轨迹分类框架中包括两个阶段，分别是训练部分和测试部分：

1) 训练阶段

A. 轨迹获取：采用 KLT 跟踪器从连续的视频帧中获取运动目标轨迹；

B. 轨迹预处理与 Multi-CTM 类标初始化：用第三章提出的 Global MRF 对原始的片段轨迹进行轨迹连接，片段轨迹连接后的每个完整轨迹树都包含一个进出口标号，用进出口标号对训练数据的类标进行初始化；

C. 基于步骤 B 中轨迹的预处理，对预处理后的完整轨迹树建立 BOW 字典，同时，进行轨迹 BOW 特征表达；

D. 将 BOW 表达后的轨迹数据和步骤 B 中初始化的数据类标作为 Multi-CTM 输入进行模型训练

2) 测试阶段

A. 将测试数据集片段轨迹进行 BOW 特征表达;

B. 用训练好的模型对步骤 A 中 BOW 表达的轨迹数据进行轨迹的类别推断。

4.4 实验结果与分析

本节给出了基于 Multi-CTM 集群目标轨迹分类的实验结果, 实验数据来自两个数据集, 分别是纽约火车站场景^[4]和交通场景^[5]。实验分为训练阶段和测试阶段, 纽约火车站场景数据集和交通场景数据集中, 用于模型训练的完整轨迹树分别为 3000 和 1050, 用于测试的片段轨迹分别为 1200 和 1000。

基于生成扩展的思想, 提出了 Multi-CTM 模型完成集群场景运动轨迹基于语义的中层特征学习和未知轨迹类别的推断, 选用 Supervised LDA^[43]和 Class-LDA^[46]作为 Multi-CTM 的实验对比方法, 改变模型主题参数 K (K 由实验经验决定, 通常取值范围为 $[2, C]$, 其中 C 为数据集中类别的数量), 交通场景数据集运动目标轨迹分类准确率见表 4-1, 纽约火车站数据集运动目标轨迹分类准确率见表 4-2。

表 4-1 交通场景数据集运动目标轨迹分类准确率

Number of Topics	Multi-CTM	Supervised-LDA	Class LDA
2	0.924	0.320	0.927
4	0.925	0.510	0.927
6	0.927	0.488	0.927
8	0.925	0.698	0.923
10	0.927	0.757	0.927
12	0.926	0.876	0.925

表 4-2 纽约火车站数据集运动目标轨迹分类准确率

Number of Topics	Multi-CTM	Supervised LDA	Class LDA
2	0.793	0.188	0.7883
5	0.791	0.417	0.790
8	0.794	0.612	0.789
11	0.794	0.683	0.788
14	0.794	0.718	0.789
17	0.794	0.693	0.788
20	0.794	0.731	0.788
23	0.794	0.721	0.788
26	0.793	0.72	0.79
29	0.793	0.733	0.788
32	0.793	0.739	0.788

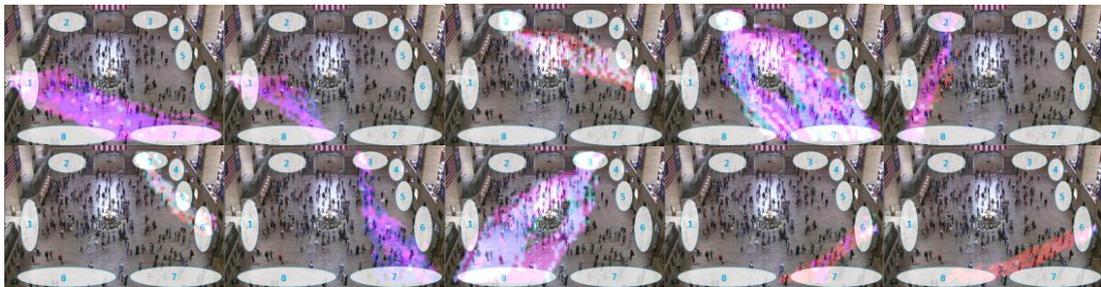
从整体上分析，纽约火车站场景属于无结构的集群场景，交通场景数据集属于有结构的集群场景，交通场景中的车辆依照遵守一定的交通规则在固定的车道上行驶，KLT 获取的原始轨迹较纽约火车站场景更有序，混杂程度和片段程度都较小。因此，交通场景数据集的分类性能整体上高于纽约火车站场景。

从单个模型的角度分析，Supervised-LDA 模型的性能随着主题个数的增大而变高，Class-LDA 和 Multi-CTM 模型的性能随着主题个数的增大变化不大。这是因为，主题模型是生成模型，不论是监督还是非监督，当主题个数非常小时，有限的主题不能够对数据背后隐藏的中层特征进行完整的表达，一个主题很可能就是多个中层特征的组合，随着主题个数的增加，生成模型虽然不能达到学习到的主题和我们期待的中层特征保持一一对应的关系，但是能够做到的是基本覆盖所有的中层特征，因此 Supervised-LDA 模型的性能会随着主题个数的增加而变高。对于 Class-LDA 和 Multi-CTM 两个模型，虽然同样是生成模型，但是，这两个模型利用不同的子模型对不同的数据进行特征学习，尽管模型的主题和

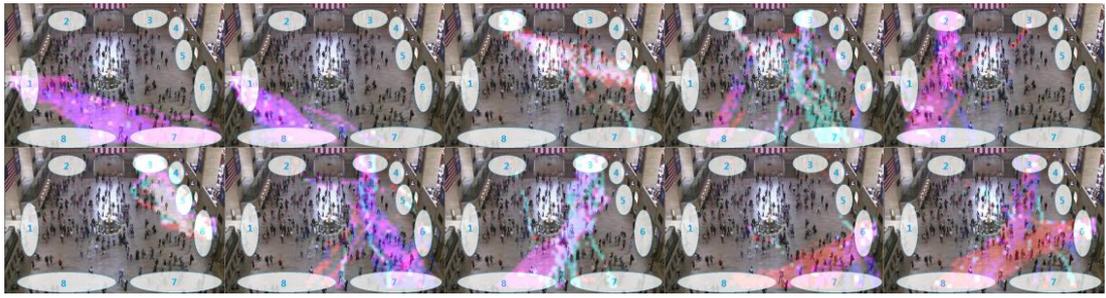
我们期待的中层特征没有一一对应，但是，每一个子模型和数据类别是意义对应的，也就是说每个子模型中不同的主题所表达的中层特征很可能非常相似。故，这两个模型的分​​类准确​​率与模型主题的个数基本无关。

从多个模型的角度分析，三个模型都是基于生成扩展的思路的扩展模型。Class-LDA 和 Multi-CTM 两个模型的思想是一致的，都是用多个子模型对不同类别的轨迹数据进行特征学习和分类，而 Supervised-LDA 是用一个模型对所有的轨迹数据进行特征学习和分类，直观的讲，用不同的子模型学习不同类别的数据的特征，那么每个子模型学习到的特征比用一个模型学习出的特征更加的准确，并且在模型和特征之间建立了一一对应的关联，也就是一个模型学习一类数据的中层特征，增大了类间距离，更有利于分类。因此，从结果上也能明显的体现出，Class-LDA 和 Multi-CTM 两种方法的分类准确​​率高于 Supervised-LDA。把 Multi-CTM 模型和 Class-LDA 模型的性能相比，在纽约火车站场景中有将近 1% 的提高，交通场景数据集中两个模型的性能基本一样，这是因为 Multi-CTM 和 Class-LDA 模型的区别在于前者的子模型是 CTM，后者的子模型是 LDA，CTM 比 LDA 的优势在于 CTM 能够利用高斯模型刻画模型内不同主题之间的关联，这样的特点在混杂的数据分类上能够增强模型的优势。另外，这样的优势在模型训练过程中学习到的中层特征有所体现，我们将在下面介绍。

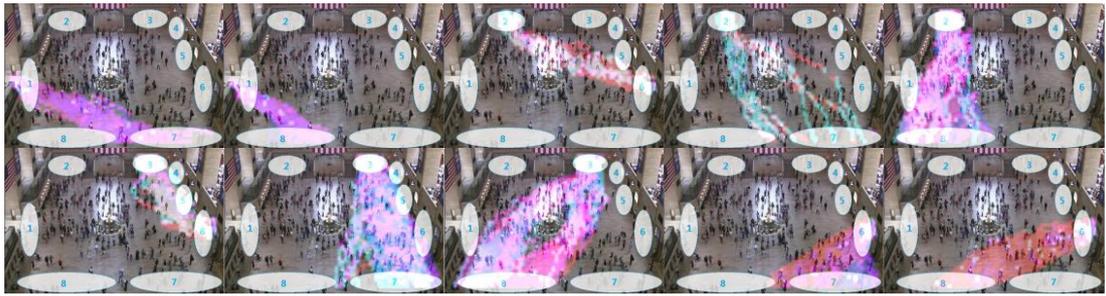
我们选取表 4-1、4-2 中三个模型性能最高点的主题个数所对应的模型训练学习得到的中层特征进行可视化，图 4-5、4-6 分别是纽约火车站数据集和交通场景数据集中层特征可视化结果，其中 (a)、(b)、(c) 是三种不同的方法，分别是 Multi-CTM、Supervised-LDA 和 Class-LDA。



(a) Multi-CTM

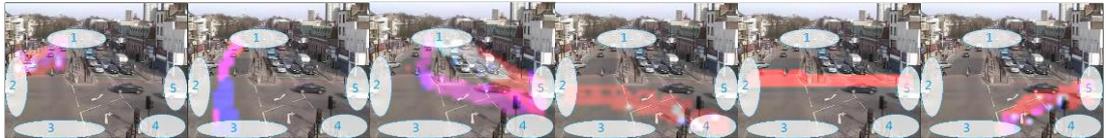


(b) Supervised-LDA

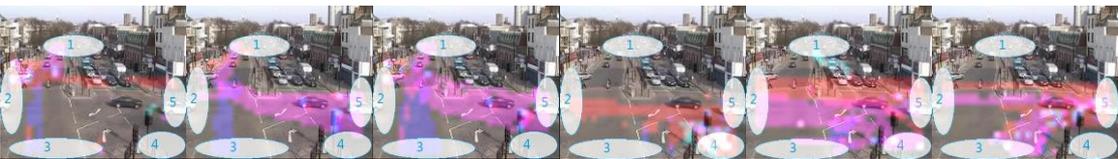


(c) Class-LDA

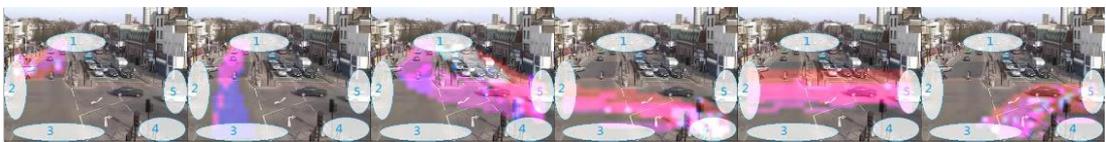
图 4-5 纽约火车站数据集中层特征可视化结果



(a) Multi-CTM



(b) Supervised-LDA



(c) Class-LDA

图 4-6 交通场景数据集中层特征可视化结果

分别对比图 4-4、4-5 中 (a)、(b) 和 (c)，显然可以发现，Multi-CTM 和 Class-LDA 模型学习的中层特征比 Supervised-LDA 的学习结果具有更强的描述力和判别力。特别的，纽约火车站场景中描述“2-7”和“3-8”的轨迹特征由于

场景中间的花坛将原本属于同一个路径的运动目标分割为两部分；交通场景中描述“1-5”的轨迹特征同样包含两种类型的轨迹，分别是右转车辆和直走车辆。模型之所以将这两类数据的特征合并成了一个是因为算法初始化模型标号的原理是相同进出口认定为同类别数据，Multi-CTM 利用子模型 CTM 中高斯分布的协方差能够发现被算法标定为同类别数据之间的关联。

4.5 本章小结

本章就 Multi-CTM 提出的思想、理论和使用进行了讨论，在该模型的基础上，着重考虑集群目标轨迹分类问题人工标定费时费力的困难，设计和实现用算法代替人工标定的集群目标轨迹分类框架，并对该框架进行了实验验证。

第五章 结论与展望

集群场景中人流密度大，拥挤程度高，轨迹分类往往面临目标轨迹不完整、轨迹混乱、人工标定代价大等挑战。本文通过对运动目标轨迹中层特征的学习和重描述实现了集群场景运动轨迹分类。

本文首先论述了课题的研究背景及意义、国内外研究现状和目前已存在的研究方法，分析了当前国内外集群场景运动目标轨迹分类存在的问题，介绍了本文的研究内容；接着在第二章详细介绍了本文研究工作中使用的相关算法和技术，为第三章和第四章的论述进行了铺垫；第三章介绍了基于BCTM集群目标轨迹分类框架的设计，详细论述了目标轨迹获取及底层特征表达；针对集群场景目标轨迹片段性高的问题，提出Global MRF模型对原始轨迹进行预处理，采用BCTM对轨迹进行中层特征提取、重描述后利用多类SVM分类器设计合理的分类流程对轨迹进行分类；第四章介绍了基于Multi-CTM集群目标轨迹分类框架的设计，借鉴生成扩展模型的思想，提出Multi-CTM，充分考虑集群场景轨迹分类问题中人工标定代价大的现实困难，详细论述了弱监督分类框架的具体实现。在上述两部分研究工作中，总结本文的主要贡献如下：

1) 针对集群监控场景中运动目标轨迹片段化和混杂的特点，提出Global MRF模型对轨迹进行预处理，将满足时空一致性的短轨迹连接成能够反映一定的运动行为模式的轨迹树。Global MRF模型克服了MRF模型局部优化的问题，并在轨迹混乱程度较高的纽约火车站数据集进行了实验，有效的验证了Global MRF模型的有效性。

2) 在集群行为理论支撑下，为了有效的利用集群场景中“从众”的运动模式进行轨迹分类，采用主题模型挖掘轨迹数据背后隐藏的基于语义的中层特征。借鉴贝叶斯模型判别式扩展的思想，设计并实现了基于BCTM和多类SVM的集群目标轨迹分类。具体的讲，采用BCTM学习运动目标轨迹的中层特征并在此基础上对目标轨迹进行重描述，利用多类SVM分类器，设计合理的分类流程对

重描述后的目标轨迹进行分类。该轨迹分类方法在两个公开监控视频数据集上进行有效的实验验证。

3) 结合贝叶斯模型生成式扩展的思想和Class LDA模型的思想,提出Multi-CTM模型对预处理后的轨迹进行中层特征学习。模型以数据样本的类别标号作为输出,通过对训练样本数据中层特征的学习,对测试样本进行类别的推断。在此模型的基础上,进一步设计并实现了基于Multi-CTM的集群目标轨迹分类方法。在具体的实现过程中,提出使用Global MRF初始化模型的类别标号,实现了弱监督的轨迹分类框架,减少了人工标定带来的代价。同样的,该分类方法在两个公开监控视频数据集上得到了有效的验证。

虽然本文提出的方法能够基本达到轨迹分类的目的,但是因为时间关系,仍然存在以下几个不足之处:

1) 基于多类SVM与BCTM集群目标轨迹分类的研究工作中,该方法有效的提高了轨迹分类的准确率,但是,在现实的集群场景智能监控中,人工标定带来代价太高,该方法在实际的使用过程中还是需要投入大量的人力;

2) 基于Multi-CTM集群目标轨迹分类的研究工作中,用算法代替人工标定工作,在大大减少人工标定的代价的同时,标定的精准性和分类的准确性都有待提高;

3) 本文提出的两种轨迹分类方法仅仅是一个离线的轨迹分析方法,距提供实时智能监控技术还有一定的距离。

参考文献

- [1] 人民网. 近两年全世界恐怖袭击案[Z]. 2015-1-18.
<http://world.people.com.cn/n/2015/0108/c14549-26351560.html>
- [2] 崇信毅, 基于视频监控数据的人群行为分析[D]. 北京: 北京交通大学, 2012.
- [3] 李秀秀, 基于目标运动特征学习的行为分析[D]. 西安: 西北工业大学, 2007.
- [4] Tomasi C., Kanade T.. Shape and Motion from Image Streams:a Factorization Method—Part 3 Detection and Tracking of Point Features[R]. *Pittsburgh: School of Computer Science, Carnegie Mellon University*. 1991.
- [5] Zhou B., Wang X., Tang X.. Random field topic model for semantic region analysis in crowded scenes from tracklets[C]. *In: Proceeding of IEEE conference on computer vision and pattern recognition*, 2011: 3441-3448.
- [6] Loy C. C., Hospedales T. M., Xiang T., et al. Stream-based joint exploration-exploitation active learning[C]. *In: Proceeding of IEEE conference on computer vision and pattern recognition*, 2012: 1560-1567.
- [7] Morris B.T., Trivedi M.M.. A survey of vision-based trajectory learning and analysis for surveillance[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2008, 18(8):1114-1127.
- [8] Hu W., Xiao X, Fu Z., et al. A system for learning statistical motion patterns[J], *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(9):1450-1464.
- [9] Morris B.T., Trivedi M.M.. Learning and classification of trajectories in dynamic scenes: A general framework for live video analysis[C]. *In: Proceeding of IEEE conferecne on advanced video and signal based surveillance*, 2008: 154-161.
- [10] Biliotti D., Antonini G., Thiran J.P.. Multi-layer hierarchical clustering of pedestrian trajectories for automatic counting of people in video sequences[C]. *In: Proceeding of IEEE workshops on application of computer vision*, 2005: 50-57.
- [11] Bashir F.I, Khokhar A.A., Schonfeld D.. Object trajectory-based activity classification and recognition using hidden markov models [J]. *IEEE Transactions on Image Processing*, 2007, 16(7): 1912-1919.
- [12] Wang X., Ma X., Grimson E.. Unsupervised activity perception by hierarchical bayesian models[C]. *In: Proceeding of IEEE conference on computer vision and pattern recognition*, 2007: 1-8.
- [13] Wold S., Esbensen K., Geladi P.. Principal component analysis[J]. *Chemometrics and intelligent laboratory systems*, 1987, 2(1): 37-52.
- [14] Abdi H., Williams L.J.. Principal component analysis[J]. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010, 2(4): 433-459.
- [15] Zhang Z., Huang K., Tan T.. Comparison of similarity measures for trajectory clustering in

- outdoor surveillance scenes[C]. *In: Proceeding of IEEE international conference on pattern recognition*, 2006, 3: 1135-1138.
- [16] Fu Z., Hu W., Tan T.. Similarity based vehicle trajectory clustering and anomaly detection[C]. *In: Proceeding of IEEE international conference on image processing*, 2005, 2: II-602-5.
- [17] Morris B., Trivedi M.. Learning trajectory patterns by clustering: Experimental studies and comparative evaluation[C]. *In proceeding of IEEE conference on computer vision and pattern recognition*, 2009: 312-319.
- [18] Hu W., Xie D., Fu Z., et al. Semantic-based surveillance video retrieval[J]. *IEEE Transaction on Image Processing*, 2007, 16(4): 1168-1181.
- [19] Keogh E.J., Pazzani M.J.. Scaling up dynamic time warping for datamining applications[C]. *In: Proceeding of international conference on knowledge discovery and data mining*, 2000: 285-289.
- [20] Vlachos M., Kollios G., Gunopulos D., Discovering similar multidimensional trajectories[C]. *In: Proceeding of IEEE conference on data engineering*, 2002: 673-684.
- [21] 曲琳, 周凡, 陈耀武. 基于 Hausdorff 距离的视觉监控轨迹分类算法[J]. 吉林大学学报: 工学版, 2009(6): 1618-1624.
- [22] 魏龙翔, 何小海, 滕奇志. 结合 Hausdorff 距离和最长公共子序列的轨迹分类[J]. 电子与信息学报, 2013, 35(4): 784-790.
- [23] Müller M. Dynamic time warping[J]. *Information retrieval for music and motion*, 2007: 69-84.
- [24] Bashir F., Qu W., Khokhar A., et al. HMM-based motion recognition system using segmented PCA[C]. *In: Proceeding of IEEE international conference on image processing*, 2005, 3: III-1288-91.
- [25] Zou J., Chen X., Wei P., et al. A belief based correlated topic model for semantic region analysis in far-field video surveillance systems[C]. *In: Proceeding of Springer International Publishing of advances in multimedia information processing-PCM*, 2013: 779-790.
- [26] Zou J., Ye Q., Cui Y., et al. A belief based correlated topic model for trajectory clustering in crowded video scenes[C]. *In: Proceeding of international conference on pattern recognition*, 2014: 2543-2548.
- [27] Chen X, Ye Q, Zou J, et al. Visual trajectory analysis via replicated softmax-based models[J]. *Signal, Image and Video Processing*, 2014, 8(1): 183-190.
- [28] Rodriguez M., Ali S., Kanade T.. Tracking in unstructured crowded scenes[C]. *In: Proceeding of IEEE international conference on computer vision and pattern recognition*, 2009: 1389-1396.
- [29] Moussaid M., Garnier S., Theraulaz G., et al. Collective information processing and pattern formation in swarms, flocks, and crowds[J]. *Topics in Cognitive Science*, 2009, 1(3): 469-497.
- [30] Zhou B., Wang X., Tang X.. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents[C]. *In: Proceeding of IEEE conference on computer vision and pattern recognition*, 2012: 2871-2878.
- [31] Li F., Perona P., A bayesian hierarchical model for learning natural scene categories[C]. *In: Proceeding of IEEE conference on computer vision and pattern recognition*, 2005, 2: 524-531.

- [32] Blei D.M., Ng A.Y., Jordan M.I.. Latent dirichlet allocation[J]. *The Journal of machine Learning research*, 2003, 3: 993-1022.
- [33] Hofmann T.. Probabilistic latent semantic indexing[C]. *In: Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval*, 1999: 50-57.
- [34]和讯科技. 百度联手微软必应围攻谷歌[Z]. 2011-7-7.
<http://www.p5w.net/news/cjxw/201107/t3699806.htm>.
- [35] Blei D. M., Lafferty J. D.. A correlated topic model of science[J]. *The Annals of Applied Statistics*, 2007: 17-35.
- [36] Crammer K., Singer Y.. On the algorithmic implementation of multiclass kernel-based vector machines[J]. *The Journal of Machine Learning Research*, 2002, 2: 265-292.
- [37] Cortes C., Vapnik V.. Support-vector networks[J]. *Machine learning*, 1995, 20(3): 273-297.
- [38] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012: 95-134.
- [39] Hsu C. W., Lin C. J.. A comparison of methods for multiclass support vector machines[J]. *IEEE transaction on neural networks*, 2002, 13(2): 415-425.
- [40] Dixit M., Rasiwasia N., Vasconcelos N.. Class-specific simplex-latent dirichlet allocation for image classification[C]. *In: Proceeding of IEEE international conference on computer vision*, 2013: 2672-2679.
- [41] Rasiwasia N., Vasconcelos N.. Latent dirichlet allocation models for image classification[J]. *IEEE transaction on pattern analysis and machine intelligence*, 2013, 35(11): 2665-2679.
- [42] Bosch A., Zisserman A., Muoz X.. Scene classification using a hybrid generative or discriminative approach[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2008, 30(4): 712-727.
- [43] Mcauliffe J. D., Blei D. M.. Supervised topic models[C]. *In: Proceeding of advances in neural information processing systems*. 2008: 121-128.
- [44] Putthividhya D., Attias H.T., Nagarajan S.S.. Supervised topic model for automatic image annotation[C]. *In: Proceeding of IEEE international conference on acoustics speech and signal processing*, 2010: 1894-1897.
- [45] Ramage D., Hall D., Nallapati R., et al. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora[C]. *In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: association for computational linguistics*, 2009: 248-256.
- [46] Quelhas P., Monay F., Odobez J.M., et al.. A thousand words in a scene[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2007, 29(9): 1575-1589.
- [47] Zhao B., Zhong Y., Zhang L.. Scene classification via latent Dirichlet allocation using a hybrid generative/discriminative strategy for high spatial resolution remote sensing imagery[J]. *Remote Sensing Letters*, 2013, 4(12): 1204-1213.
- [48] Russell B.C., Torralba A., Murphy K.P., et al. LabelMe: a database and web-based tool for image annotation[J]. *International journal of computer vision*, 2008, 77(1-3): 157-173.

附录：主要符号对照表

BCTM	基于置信关联主题模型 (Belief based Correlated Topic Model)
BOW	词包 (Bag-of-Words)
Class-LDA	关联主题模型 (Correlated Topic Model)
CTM	关联主题模型 (Correlated Topic Model)
DTW	动态时间整合 (Dynamic Time Warping)
FCM	模糊 C 均值 (Fuzzy C-means)
Global-MRF	全局马尔科夫随机场 (Global Markov Random field)
KLT	关键点跟踪器 (Kanade-Lucas-Tomasi)
LCSS	最长公共子序列 (Longest Common Subsequence)
LDA	隐狄利克雷分布 (Latent Dirichlet Allocation)
多类 SVM	多类支持向量机 (Multiclass Support Vector Machine)
Multi-CTM	多关联主题模型 (Multi-Correlated Topic Model)
PCA	主成分分析 (Principal Component Analysis)
PLSA	概率潜在语义分析 (Probabilistic Latent Semantic Analysis)
RFTM	条件随机场主题模型 (Random Field Topic Model)
SIFT	尺度不变特征转换 (Scale-invariant feature transform)
Supervised LDA	监督隐狄利克雷分布 (Supervised Latent Dirichlet Allocation)

致 谢

在中国科学院大学攻读硕士学位三年的学习生活中，我经历了诸多坎坷，付出了艰辛的努力，同时也得到了很大的收获。在毕业论文完成之际，由衷的感谢这三年曾经给予我无数帮助的老师、同学、朋友和家人。

本课题的研究工作及学位论文是在焦建彬教授和叶齐祥副教授的悉心指导下完成的。首先，我要感谢导师叶齐祥副教授在我攻读硕士学位期间从生活和学习各个方面给予的关怀和指导，以及在我论文的选题、框架的设计、论文的撰写和修改中倾注的大量心血。感谢焦建彬教授在三年的学习中对我每一点不足和进步的及时指正和鼓励，以及生活中给予的亲切关怀和无私帮助。两位恩师严谨的治学态度和平易近人的学者风范深深地影响着我，他们诲人不倦的精神让我感动，这些言传身教将使我受益终身。

其次，感谢韩振军老师、邹佳凌师兄和魏朋旭师姐在我的理论学习和课题研究过程中提供的耐心的指导和帮助。三年来，他们渊博的专业知识，扎实的动手能力和自强不息的学习精神深深影响地影响着我，从中我也收获了一些令我受益匪浅的宝贵知识。

衷心的感谢我的师兄师姐和同学们，感谢学习生活中大家的相互帮助，相互鼓励，一起渡过了三年的快乐时光，结下了深厚的友谊。特别是好友杨倩，庞丽金，孔微，刘波，王忻雷。三年中，他们曾和我一起经历成长中的喜悦和困惑，这段美好的回忆在我人生当中永不退色。

特别要感谢我的家人，感谢他们多年来一直给我的爱和支持，永远是最坚强的后盾和避风的港湾，愿他们一生平安健康。

感谢参加开题及中期评阅的各位老师和专家们，他们曾经利用丰富的经验来无私地帮助我把握论文方向和研究进度。

最后，感谢参加论文评审和答辩的各位老师。

崔妍婷
2015年4月

个人简介、在学期间发表的论文与研究成果

姓名：崔妍婷 性别：女 出生日期：1990.03.25 政治面貌：中共党员

教育经历

- 2008年9月至2012年7月 青岛理工大学 计算机科学与技术 学士
- 2012年9月至2015年7月 中国科学院大学 计算机技术 硕士

已发表（录用）论文

1. Jialing Zou, Qixing Ye, **Yanting Cui**, David Doermann, Jianbin Jiao. A Belief based Correlated Topic Model for Trajectory Clustering in Crowded Video Scenes International Conference on Pattern Recognition (ICPR), 2014 : 2543-2548
2. Jialing Zou, **Yanting Cui**, Fang Wan, Qixiang Ye, Jianbin Jiao. A Cluster Specific Latent Dirichlet Allocation Model for Trajectory Clustering in Surveillance Videos. International Conference on Image Processing (ICIP), 2014.
3. Xiaogang Chen, Qixiang Ye, Jialing Zou, Ce Li, **Yanting Cui**, Jianbin Jiao. Visual Trajectory Analysis via Replicated Softmax based Model. Signal, Image and Video Processing, 2014 : 8(1), 183-190.

申请专利

焦建彬, **崔妍婷**, 邹佳凌, 万方, 叶齐祥, 韩振军, 基于集群轨迹分类的集群场景智能监控方法和系统, 201510100197.0, 中国发明型专利