

基于模态间哈希码损失的检索性能优化路径

王萌 梁璐 卢泽涛

广西民族大学相思湖学院, 广西南宁, 530000;

摘要: 哈希码是一种算法, 用于将任意长度的数据映射为固定长度的值。在图文检索场景中, 跨模态哈希码常用于在比特空间内实现近似邻近搜索。训练阶段的模态间哈希码损失却与检索指标之间缺乏清晰对应。本文以模态间哈希码损失为主线, 将跨模态检索性能问题拆解到表征、码字和损失三个层面。在图文特征学习与哈希空间构建的基础上, 为模态间哈希码损失设计语义相似度建模、结构分解和权重调度方案。同时利用损失反馈, 对特征表示和哈希码字进行重构与重排, 使训练目标与实际检索过程逐步一致。

关键词: 跨模态; 哈希码损失; 图文特征

DOI: 10.64216/3104-9702.25.08.024

引言

大规模图文数据在内容平台、搜索服务和多媒体档案中持续累积, 用户往往以图找文或以文找图。跨模态哈希利用紧凑二值码支撑高效索引, 已经在实际系统中占据关键位置, 不过很多实现偏重网络结构堆叠, 对训练损失与检索行为之间的关系关注不足。模态间哈希码损失决定语义相似度在比特空间中的投射形态, 也影响索引结构能否在高并发访问下保持稳定。若缺乏有意识的设计与调节, 损失项很难一并兼顾精度和效率。本文围绕模态间哈希码损失, 重新审视图文特征学习与哈希空间构建过程, 提出面向表征层、码字层和损失层的多级优化路径。

1 跨模态哈希检索与模态间哈希码损失的理论基础

跨模态哈希检索的理论基础根植于大规模多媒体数据共享表示的需求。在图像与文本耦合的检索场景中, 系统需要在离散比特空间中刻画跨模态语义相似度, 使哈希距离与语义邻近关系保持稳定的对应。哈希码损失是指在哈希学习中, 由于量化误差和相似性保留的约束, 导致生成的哈希码与原始数据之间的差异。

从信息编码角度看, 哈希码承担压缩表示与快速索引的双重职责, 模态间损失项约束不同模态在语义锚点处形成可分的比特模式, 以减轻模态偏移带来的距离畸变^[1]。结合统计学习视角, 该类损失隐含对跨模态条件分布的重塑: 负似然形式拉开相似与不相似配对的概率边界, 多尺度投影匹配项保持局部结构的一致性, 对抗投影学习抑制采样偏差与标注噪声引起的判别失衡, 由此构成的损失体系既为哈希网络提供可微的优化信号, 也决定比特分布、索引结构以及在高并发检索中的鲁棒性表现。

2 图文特征学习与哈希空间构建

2.1 图文特征学习

特征学习 (Feature Learning) 是机器学习和深度学习中的一个核心概念, 其目的是通过算法自动从数据中学习有效的特征表示, 而不是依赖人工设计特征。特征学习的目标是让模型从原始数据中提取和表示有意义的信息, 以便在分类、回归、生成等任务中获得更高的性能。目前, 图文特征学习面临的核心矛盾在于模态差异与语义共性的并存。视觉模态在像素层面呈现局部相关的栅格结构, 语言模态则以离散符号序列承载语义与语法约束, 两者在统计分布和噪声类型上差异明显。面向跨模态哈希检索, 需要在预处理阶段把这类差异压缩在可控范围内, 例如在图像侧进行亮度归一、尺度重采样以及结构增强, 在文本侧结合中文分词、子词编码与停用词剔除, 使原始数据进入特征提取网络时不再被低价值细节干扰。表征学习阶段更强调语义层级的构建, 深层视觉网络逐步聚合边缘、纹理与区域布局, 预训练语言模型则在上下文依存关系中抽取实体、属性与事件模式。现实数据集普遍存在标注稀疏、语义细粒度多样等问题, 因此图文特征学习不宜停留在简单配对记忆, 而应在局部邻域结构中保留类内差异和跨标签共现关系, 为后续跨模态哈希映射提供兼具判别性与可压缩性的基础表示^[2]。

2.2 哈希空间构建

哈希空间构建关注的是如何把连续联合表征嵌入到离散比特空间, 使相似关系在代价有限的编码长度内得到最大程度保留。构建过程常以一组可训练的投影单元为起点, 这些线性或弱非线性映射把公共特征压缩到中间维度, 再在接近二值边界的区域设置平滑过渡机制, 减少硬阈值带来的梯度断裂和数值不稳定。与传统度量学习只关注欧氏距离不同, 哈希空间设计需要直接面向

汉明距离的几何结构,考虑哈希球覆盖范围、码字间最小间隔以及不同类别在码空间中的分布疏密关系^[3]。编码长度与数据规模、索引结构紧密相关,过短会导致碰撞频繁,过长又会显著增加存储与比对开销,因此实际系统往往在可用存储、检索延迟和并发访问量之间寻找折中。

3 模态间哈希码损失的建模与分解

3.1 模态间语义相似度建模

模态间哈希码损失的建模起点在于对跨模态语义关系的精细刻画,而不是简单依赖二元相似标签。面向图文检索,需要在样本层面区分强相关、弱相关与语义无关的配对,使哈希距离变化对应这种细粒度差异。多标签标注构建加权相似度矩阵,把标签交集规模与类别相关性并入权重,使损失在更新时重点修正语义边界模糊的配对。负似然对数形式适合描述这类加权相似度,因为它能够把离散相似标签映射为连续概率约束,并在梯度空间中形成稳定的拉近与拉远信号^[4]。相较仅依赖硬阈值的距离约束,这种概率化表达更贴近检索任务中的排序需求,在高相似度配对之间形成更细区分,在语义差异显著的配对上保持足够间隔,为后续损失结构分解奠定可操作的量化基础。

3.2 哈希码损失结构的分层设计

模态间哈希码损失的结构设计并非简单叠加约束项,而是在语义对齐、量化误差与比特统计性质之间建立清晰分工。语义对齐项直接约束跨模态配对的哈希距离,使同一语义簇的图文编码在比特模式上保持重叠,无关样本维持间隔。量化相关项围绕连续输出向离散码字的过渡构造惩罚,限制特征投影停留在比特边界附近的中间区域,减弱检索排序对微小扰动的敏感度。比特分布正则从整体统计角度修正哈希空间,一方面均衡各比特取值频率以缓解碰撞,另一方面削减比特冗余,使有效容量与检索规模相匹配。将这三类损失剥离为层次分明的子结构,便于在训练过程中分别监测梯度贡献与数值尺度,避免某一约束长期主导更新而压制模态间相似度建模的作用。

3.3 损失项权衡与优化目标

模态间哈希码损失在训练中的表现取决于各损失项之间的权衡,因此在优化目标层面需要对权重分配给出明确规则。检索系统通常同时关注检索精度与效率,语义对齐项偏向提升排名质量,量化及正则项关系到码字稳定性和索引表现。静态设定固定权重难以适应训练阶段侧重点的迁移,实践中依据信号强度、收敛速度或验证集指标,对损失权重实行自适应调整,使模型在早期更侧重构建可靠的模态间距离结构,在中后期逐步强

化比特分布与泛化能力。针对难样本和边界样本,引入重加权机制,提高这些配对在损失中的占比,缓解数据不平衡导致的哈希空间偏斜。经由这类权衡策略的嵌入,模态间哈希码损失呈现出可调节特性,为检索性能的持续提升提供更细致的控制手段。

4 基于模态间哈希码损失的检索性能优化路径

为了便于在工程实践中落实基于模态间哈希码损失的检索性能优化,可将整体工作抽象为一条由统一管理驱动的分层路径,可见图1。

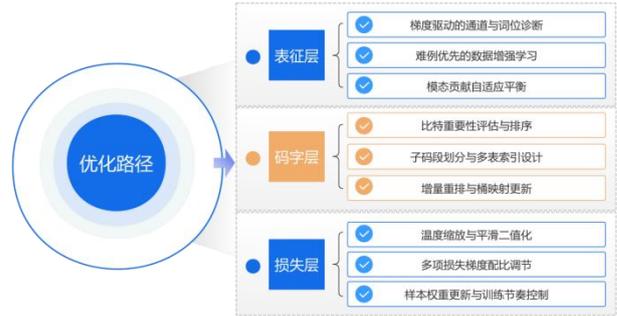


图1 基于模态间哈希码损失的检索性能优化路径

4.1 表征层：利用损失反馈的跨模态特征重构

表征层的优化以模态间哈希码损失的细粒度反馈为核心信号,在训练过程中对图像特征张量和文本隐状态分别记录梯度范数分布,将高损失样本中梯度异常集中的通道或词位标记为不稳定区域。网络结构上,可在主干编码器之后插入轻量的特征校准模块,例如针对视觉模态设置通道重标定层和局部区域门控,对文本模态引入子词级重加权单元,而这些附加参数只接受模态间损失项的反向更新,使校准行为紧贴跨模态对齐需求。训练迭代进入稳定阶段时,可以构造基于损失排序的难例队列,把哈希距离偏离语义关系的图文配对集中组成小批量,将其送入带有更强数据扰动的增强支路,对视觉侧采用随机裁剪、颜色扰动,对文本侧采用同义表达替换或短语重组,促使表征在局部扰动下保持跨模态邻域关系不改变。为防止强势模态长期主导公共空间,还需要定期统计不同模态在损失中的贡献比例,对梯度进行模态级缩放,在保持总体收敛速度的前提下提升弱势模态在共享子空间中的表达权重,形成由损失驱动的特征重构闭环。

4.2 码字层：面向检索路径的哈希结构重排

码字层的优化不再调整特征编码网络,而是围绕已有哈希输出构建更符合检索路径的离散结构。训练阶段完成后,可以在全量训练集上收集连续哈希输出和对应二值码字,统计各比特位的取值频率以及与跨模态相似度矩阵的相关系数,据此为比特位建立重要性排序,将信息贡献低且噪声敏感的维度聚集在尾部,优先分配高

重要性比特给实际检索索引。面向大规模检索场景,可将完整码字划分为若干子段,每个子段对应一张子索引表,同时结合历史模态间哈希码损失中误检样本的分布,确定各子段在检索顺序中的优先级,使高判别性的子段在候选召回阶段优先参与匹配^[9]。在线查询时,为每个哈希查询动态设定探测半径和候选桶数量,这一参数不采用固定配置,而是依据离线统计得到的“损失敏感比特集合”决定优先翻转的位,沿预定义的比特翻转序列在汉明邻域内展开检索,从而在不改变网络参数的前提下调整实际搜索路径。对持续增长的数据流,在滑动时间窗口内维护比特分布估计,当偏移超过阈值时触发轻量级重排,只改写索引结构与桶映射,避免频繁整网重训。

4.3 损失层:面向训练过程的多阶段目标调度

损失层的优化路径侧重训练过程的时序组织与权重调度,使模态间哈希码损失在不同阶段承担不同角色。初始训练阶段采用温度缩放策略,将哈希输出视作平滑的实值响应,设定较高温度以拉宽决策边界,使梯度在二值边界附近保持稳定,再根据验证集上的检索指标逐步降低温度,促使输出靠近离散码字。在多项损失并存的情况下,可以在每个小批量中统计各子项对参数更新的梯度贡献,构造归一化系数,使任一子项的梯度比例不长期压制其他部分,并在训练中后期将权重向与实际检索指标更相关的项倾斜。样本层面,为减轻易分类样本的干扰,可维护一个基于历史损失的指数滑动平均,对长期处于低损失区间的配对降低权重,把更新资源集中在跨模态距离与语义关系不一致的困难配对上,从而提升损失信号的有效密度。在实现层面,还需对模态间损失对应的梯度单独设置裁剪阈值与学习率缩放,避免极端难例引发训练不稳定,同时在阶段切换节点固定一小段迭代,仅更新与哈希输出直接相连的层,使损失层的调度与网络各部分的收敛节奏保持协调。

5 基于模态间哈希码损失的检索性能优化路径评价

基于模态间哈希码损失的优化路径需要在统一评价框架下检验效果,否则无法判断各层改造是否真正作用于检索链路。离线阶段可在固定数据集和相同哈希长度条件下,构建多组对照实验:保留原始模型作为基线,分别启用表征校准、码字重排和损失调度模块,记录每一步引入后在 MAP、NDCG、Recall@K 等指标上的增幅,并同步采集索引构建时间、单次查询延迟以及内存占用。日志侧写入更细粒度信息,例如比特翻转频率、梯度方差、难例集合的收敛步数,由此形成针对不同路

径单元的诊断矩阵。评价过程中还要在多次随机初始化下重复训练,统计指标分布的方差和极值,检验优化路径对收敛稳定性的影响。在线系统可选取一段真实检索流量,设计哈希索引层面的 A/B 实验,把新路径产生的码字和旧方案解耦部署,比较点击率、有效停留时长以及用户跳出率等行为信号,且结合查询日志拆分热门查询、长尾查询和新类别查询,分析各类场景下的收益差异。若离线与在线结果在趋势上保持一致,且在资源开销满足业务约束的前提下保持稳定增益,则可以认定该优化路径在工程实践中具备可推广价值。

6 结语

这一工作围绕模态间哈希码损失构建检索性能优化路径,可以把跨模态哈希从孤立的网络设计转向以目标驱动的整体工程。本文在图文特征学习与哈希空间构建的基础上,梳理损失在语义建模、结构拆分和权重配置中的作用,还据此形成表征层校准、码字层重排以及损失层调度等协同环节。这些路径把训练信号与实际检索操作联系起来,使语义对齐、比特分布和索引结构在统一框架内得到调节。后续工作有望在视频、音频等更复杂模态组合中推广这套分层设计,在增量学习场景中探索与索引更新协同的训练机制。除此以外,还可在隐私约束条件下考察哈希码分布与安全性的平衡,从工程和理论两个方向推动跨模态哈希检索迈向更大规模和更高可靠性。

参考文献

- [1] 殷崧祚,李博涵,王萌,等.基于细粒度特征融合的部分多模态哈希[J].软件学报,2024,35(03):1074-1089.
- [2] 朱杰.基于文本引导对抗哈希的跨模态检索方法[J].计算机应用研究,2022,39(02):628-632.
- [3] 廖列法,李志明,张赛赛.基于深度残差网络的迭代量化哈希图像检索方法[J].计算机应用,2022,42(09):2845-2852.
- [4] 马宾,王一利,徐健,等.基于双向生成对抗网络的图像感知哈希算法[J].电子学报,2023,51(05):1405-1412.
- [5] 余仁杰,姜军,张少华,等.融合 YOLOv8n 与 BERT 多模态特征提取的多标签图文分类研究[J].软件导刊,2025,24(12):197-204.

广西民族大学相思湖学院 2024 年度校级科研项目《基于特征学习的图文数据跨模态哈希检索研究》(2024XJKY48)