

# CF 患者肺组织分类——基于相关反馈的监督核哈希方法

申华磊, 邱鹏

(河南师范大学 计算机与信息工程学院, 河南 新乡 453007)

**摘要:** 现有的哈希方法用于 CF 患者肺组织分类时没有从正负反馈样本挖掘判别信息, 分类精度不高. 为此, 提出一种基于相关反馈的监督核哈希方法. 首先, 对肺组织进行监督核哈希学习, 得到初始哈希函数; 其次, 使用该初始哈希函数对肺组织进行哈希编码和分类, 并得到正负反馈样本; 接着, 基于正负反馈样本构建新的哈希函数; 最后, 使用新构建的哈希函数对肺组织再次进行哈希编码和分类. 实验结果表明, 同现有方法相比, 所提出的方法显著提高了 CF 患者肺组织的分类精度.

**关键词:** 相关反馈; 监督核哈希; 肺组织分类; CT 影像

**中图分类号:** TP391

**文献标志码:** A

肺部感染是囊性纤维化(Cystic Fibrosis, CF)患者死亡的主要病因之一. 定量分析 CF 患者肺组织的感染程度, 制定个性化治疗方案, 对治疗 CF 具有重要意义<sup>[1]</sup>. 当前, CF 患者肺组织分类由影像科医生手工完成. 如图 1(a) 所示, 对一幅肺部计算机断层扫描(Computed Tomography, CT)影像, 影像科医生对其中的每一块肺组织区域进行分类, 并标注为不同的颜色(见图 1(b)). 其中, 如图 1(c) 所示, 绿色、品红色、红色、橙色、黄色和蓝色分别表示对应肺组织的类别为健康(类别 1)、肺萎缩(类别 2)、支气管扩张(类别 3)、气道异常(类别 4)、黏液栓塞(类别 5)和空气残留(类别 6).

手工对肺组织分类存在两点不足. 一是医生工作负担重, 分类效率不高. 每一帧肺 CT 影像平均有 200 个区域需要医生标注; 一位患者的肺 CT 影像平均包含 300 帧, 有 60 000 个区域需要标注. 二是存在一定的误判率. 对于海量需要标注肺组织的 CT 影像, 影像科医生很难避免误诊.

使用图像分类技术辅助影像科医生对肺组织进行自动分类, 是解决上述问题的有效方法<sup>[2]</sup>. 哈希方法是有效的图像分类方法<sup>[3-7]</sup>; 该类方法使用哈希函数, 将图像视觉特征从高维空间映射到低维空间, 之后进行分类. 有代表性的哈希方法有谱哈希(Spectral Hashing, SH)<sup>[8]</sup>、局部敏感哈希(Local Sensitive Hashing, LSH)<sup>[9]</sup>、迭代量化哈希(Iteration Quantization Hashing, IQH)<sup>[10]</sup>和监督核哈希(Kernel-based Supervised Hashing, KSH)<sup>[11]</sup>等.

SH 借助谱分析(Spectral Analysis, SA)技术, 从高维特征空间学习出图像的低维特征表达. LSH 将图像从高维特征空间投影到低维特征空间时, 保留了图像特征在高维特征空间内的判别信息. IQH 使用迭代优化技术, 最小化图像在低维特征空间内的旋转矩阵误差, 从而尽可能保留图像的判别特征. KSH 通过监督学习技术构造从高维特征空间到低维特征空间的转换矩阵, 之后使用贪婪优化算法求解该矩阵.

哈希方法已开始应用于医学图像分类. Liu 等人<sup>[12]</sup>基于 IQH 的思想, 将多模态医学影像映射到低维空间; 以此为基础, 对乳腺 X 射线图像进行分类. Zhang 等人<sup>[13]</sup>使用 KSH 将医学影像的特征从高维空间映射到低维空间, 进行大规模医学影像分类. 文献<sup>[14]</sup>结合多特征学习和 IQH 的思想, 使用多核学习赋予多个特征不同权重, 得到医学图像的最优哈希编码.

**收稿日期:** 2018-07-23; **修回日期:** 2019-01-21.

**基金项目:** 国家自然科学基金(61502319); 河南省科技攻关项目(172102210337; 182102210363); 河南师范大学博士启动经费支持课题(qd16120).

**作者简介(通信作者):** 申华磊(1981-), 男, 河南西平人, 河南师范大学副教授, 博士, 研究方向为计算机辅助医学影像分析, 深度学习, E-mail: shenhualei@henannu.edu.cn.

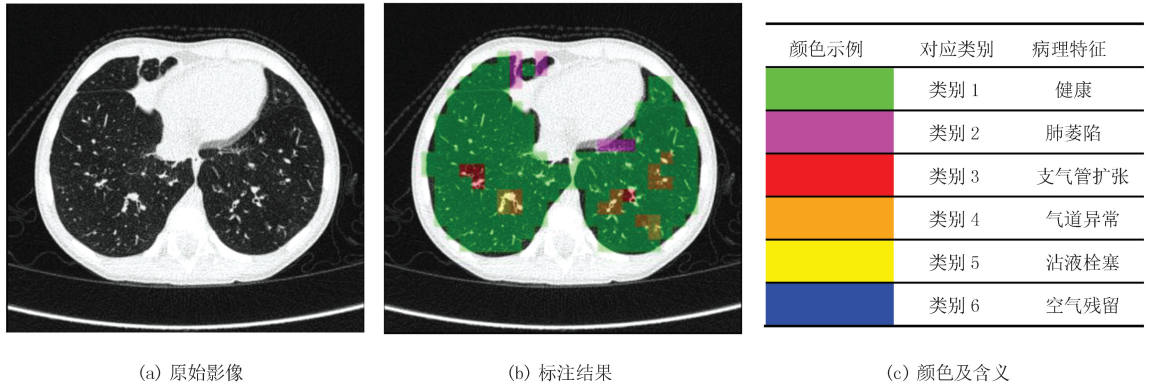


图 1 CF 患者肺 CT 影像示例

Fig.1 Example lung CT images from CF patients

在图像检索领域,相关反馈(Relevance Feedback, RF)<sup>[15]</sup>的基本思想是把图像检索结果序列分为正样本(Positive Sample, PS)和负样本(Negative Sample, NS).其中正样本为分类正确的样本序列;负样本为分类错误的样本序列.基于这些反馈样本,学习出有效区分正负样本的判别特征或分类器.之后,使用该判别特征或分类器重新进行图像分类,从而提高分类精度.

RF 已应用于医学图像分析领域,Shen 等人<sup>[16]</sup>提出一种基于子空间学习的 RF 方法,从反馈得到的正负样本学习出不同类别医学影像的判别特征,从而提高医学图像分类的精度.Huang 等人<sup>[17]</sup>从正负反馈样本学习待分类样例的判别特征,以剔除初始检索结果包含的噪声,从而提高医学图像分类的精度.Behnam 等人<sup>[18]</sup>从正反馈样本中学习出待分类样例的密度函数,以此提高多模态医学影像分类的精度.Banerjee 等人<sup>[19]</sup>通过 RF 机制融合医学影像的底层特征和语义特征,从而实现骨肿瘤影像的高效分类.

现有的哈希方法很少和 RF 技术结合.这导致对于解决如 CF 患者肺组织分类等比较复杂的医学影像分析问题时,直接使用哈希方法不能取得满意的结果.为此,本文提出一种基于相关反馈的监督核哈希方法(Relevance Feedback-based Kernel Supervised Hashing, RFKSH).RFKSH 能够从正负反馈样本有效学习出每类肺组织的判别特征,从而提高 CF 患者肺组织自动分类的精度.

本文剩余部分的结构安排如下:第 1 节介绍监督核哈希的基本方法;第 2 节阐述本文提出的 RFKSH 算法;第 3 节报告实验结果;第 4 节总结论文.

## 1 监督核哈希

为便于描述,表 1 列出本文使用的重要符号及其含义.

记  $I = \{I_1, I_2, \dots, I_N\}$  为一个包含  $N$  个样本的肺组织数据集;其对应的高维特征向量集为  $X_N = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N\}$ .定义函数集  $H(\vec{x}) = \{f_1(\vec{x}), f_2(\vec{x}), \dots, f_r(\vec{x})\}$ ,其中核函数  $f_i(\vec{x}) \in \{-1, 1\}$ , ( $1 \leq i \leq r$ ,  $\vec{x} \in X_N$ ), 为哈希函数:

$$f_i(\vec{x}) = \sum_{i=1}^m k(\vec{x}_{(i)}, \vec{x}) a_i - b. \quad (1)$$

定义  $X_m = \{\vec{x}_{(1)}, \vec{x}_{(2)}, \dots, \vec{x}_{(m)}\}$  为从  $X_N$  中随机选取的包含  $m$  个样本的集合,称其为锚点集; $k(\vec{x}_{(i)}, \vec{x}) \in \mathbf{R}^{1 \times 1}$  为核函数,  $\vec{x}_{(i)} \in X_m$ , ( $1 \leq i \leq m$ ),  $\vec{x} \in X_N$ ,  $a_i$  为系数,  $b$  为偏置.取

$$b = \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^m k(\vec{x}_{(i)}, \vec{x}_j) a_i, \quad (2)$$

则(1)式化为:

$$f_i(\vec{x}) = \sum_{i=1}^m k(\vec{x}_{(i)}, \vec{x}) a_i - \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^m k(\vec{x}_{(i)}, \vec{x}_j) a_i = \vec{a}_i^T \vec{k}(\vec{x}), \quad (3)$$

其中  $\vec{k}(\vec{x}) = [k(\vec{x}_{(1)}, \vec{x}) - \mu_1, \dots, k(\vec{x}_{(m)}, \vec{x}) - \mu_m]^T \in \mathbf{R}^{m \times 1}$ ,  $\vec{a}_i = [a_{i1}, a_{i2}, \dots, a_{im}]^T \in \mathbf{R}^{m \times 1}$ , 为系数向量,  
 $\mu_i = \sum_{j=1}^N k(\vec{x}_{(i)}, \vec{x}_j) / N, 1 \leq i \leq m$ .

定义  $\mathbf{K}_l = [\vec{k}(\vec{x}_1), \vec{k}(\vec{x}_2), \dots, \vec{k}(\vec{x}_l)]^T \in \mathbf{R}^{l \times m}$ , 其中  $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_l\} = X_l$  为包含  $l$  个样本的训练集, 得到:

$$\mathbf{H}_l \mathbf{H}_l^T = \sum_{k=1}^r \text{sgn}(K_l a_k) (\text{sgn}(K_l a_k))^T, \tag{4}$$

其中

$$\mathbf{H}_l = \begin{bmatrix} h_1(\vec{x}_1) & \cdots & h_r(\vec{x}_1) \\ \vdots & & \vdots \\ h_1(\vec{x}_l) & \cdots & h_r(\vec{x}_l) \end{bmatrix} = \text{sgn}(K_l \mathbf{A}), \tag{5}$$

$\text{sgn}(x) \in \{-1, 1\}$  为符号函数;  $\mathbf{A} \in [\vec{a}_1, \vec{a}_2, \dots, \vec{a}_r] \in \mathbf{R}^{m \times r}$  为系数矩阵.

定义符号矩阵  $\mathbf{S} \in \{-1, 1\}^{l \times l}$ :

$$S_{ij} = \begin{cases} +1, & \vec{x}_i, \vec{x}_j \text{ 属于同类,} \\ -1, & \vec{x}_i, \vec{x}_j \text{ 属于不同类,} \end{cases} \tag{6}$$

其中  $\vec{x}_i, \vec{x}_j \in X_l; 1 \leq i, j \leq l$ .

定义核监督哈希的目标函数为:

$$\arg \min_{\mathbf{H}_l} \left\| \frac{1}{r} \mathbf{H}_l \mathbf{H}_l^T - \mathbf{S} \right\|_F^2 = \arg \min_{\mathbf{A} \in \mathbf{R}^{m \times r}} \left\| \sum_{k=1}^r \text{sgn}(K_l \vec{a}_k) (\text{sgn}(K_l \vec{a}_k))^T - r \mathbf{S} \right\|_F^2 \tag{7}$$

定义矩阵  $\mathbf{Q}_{k-1} = r \mathbf{S} - \sum_{i=1}^{k-1} \text{sgn}(K_l \vec{a}_i^*) (\text{sgn}(K_l \vec{a}_i^*))^T$ , (7) 式中的系数矩阵  $\mathbf{A}$  可通过贪婪优化算法求解.

表 1 重要符号及含义

Tab.1 Important symbols and their meanings

符号	含义	符号	含义
$\mathbf{R}$	实数集	$\mathbf{S}$	符号矩阵
$I$	肺组织数据集	$\mathbf{A}$	初始系数矩阵
$N$	$I$ 的大小	$\mathbf{A}'$	引入相关反馈后的系数矩阵
$X_N$	$I$ 对应的高维特征向量集	$b$	偏置
$H(\vec{x})$	初始哈希函数集	$\vec{x}_i$	$X_i$ 的特征向量
$H'(\vec{x})$	引入 RF 后的哈希函数集	$f_i(\vec{x})$	第 $i$ 个哈希函数
$X_m$	初始锚点集	$\text{sgn}(\vec{x})$	符号函数
$X'_m$	引入 RF 后的锚点集	$r$	哈希位数
$X_l$	初始训练集	$\vec{a}_i$	第 $i$ 个哈希函数的系数向量
$X'_l$	引入 RF 后的训练集	$m$	锚点集的大小
$\bar{X}_l$	包含 $N-l$ 个样本的测试集	$l$	训练集的大小
$X_i$	$X_N$ 中的第 $i$ 个样本	$N-l$	测试集的大小

## 2 RFKSH 算法

本文提出的 RFKSH 算法如图 2 所示, 共包含以下 6 个步骤.

**步骤 1** 初始哈希函数集生成. 从  $X_N$  中分别随机选取  $m$  个和  $l$  个样本, 构建初始锚点数据集  $X_m$  和初始训练样本数据集  $X_l$ . 根据公式(6) 构建监督矩阵  $\mathbf{S}$ , 根据公式(7) 求得系数矩阵  $\mathbf{A}$ . 从而求得初始哈希函数集  $H(\vec{x})$ .

**步骤 2** 初始分类.使用初始哈希函数集  $H(\vec{x})$ ,分别求得  $X_N$  和待分类肺组织的哈希表示.对于每一个待分类样本,以其为检索样例,从  $X_N$  中检索和其最相似的前  $k$  个样本.在这  $k$  个样本中,统计每一类样本的个数.包含样本数目最多的类别,即为待分类肺组织的初始分类类别.

**步骤 3** 新的哈希函数集生成.把步骤 2 获得的  $k$  个检索结果分为正样本和负样本.以此为基础构建新的锚点数据集  $X'_m$  和训练样本数据集  $X'_l$ .根据公式(7) 求得新的系数矩阵  $A'$  和新的哈希函数集  $H'(\vec{x})$ .

**步骤 4** 获取  $X_N$  的新的哈希表示.根据步骤 3 得到的新的哈希函数集  $H'(\vec{x})$  求得  $X_N$  的新的哈希表示.

**步骤 5** 获取待分类肺组织的新的哈希表示.根据步骤 3 得到的新的哈希函数集  $H'(\vec{x})$  求得待分类肺组织的新的哈希表示.

**步骤 6** 获取最终分类结果.对于每一个待分类组织,在步骤 5 获取的新的哈希表达上,重新进行步骤 2 中的检索,得到其最终分类结果.

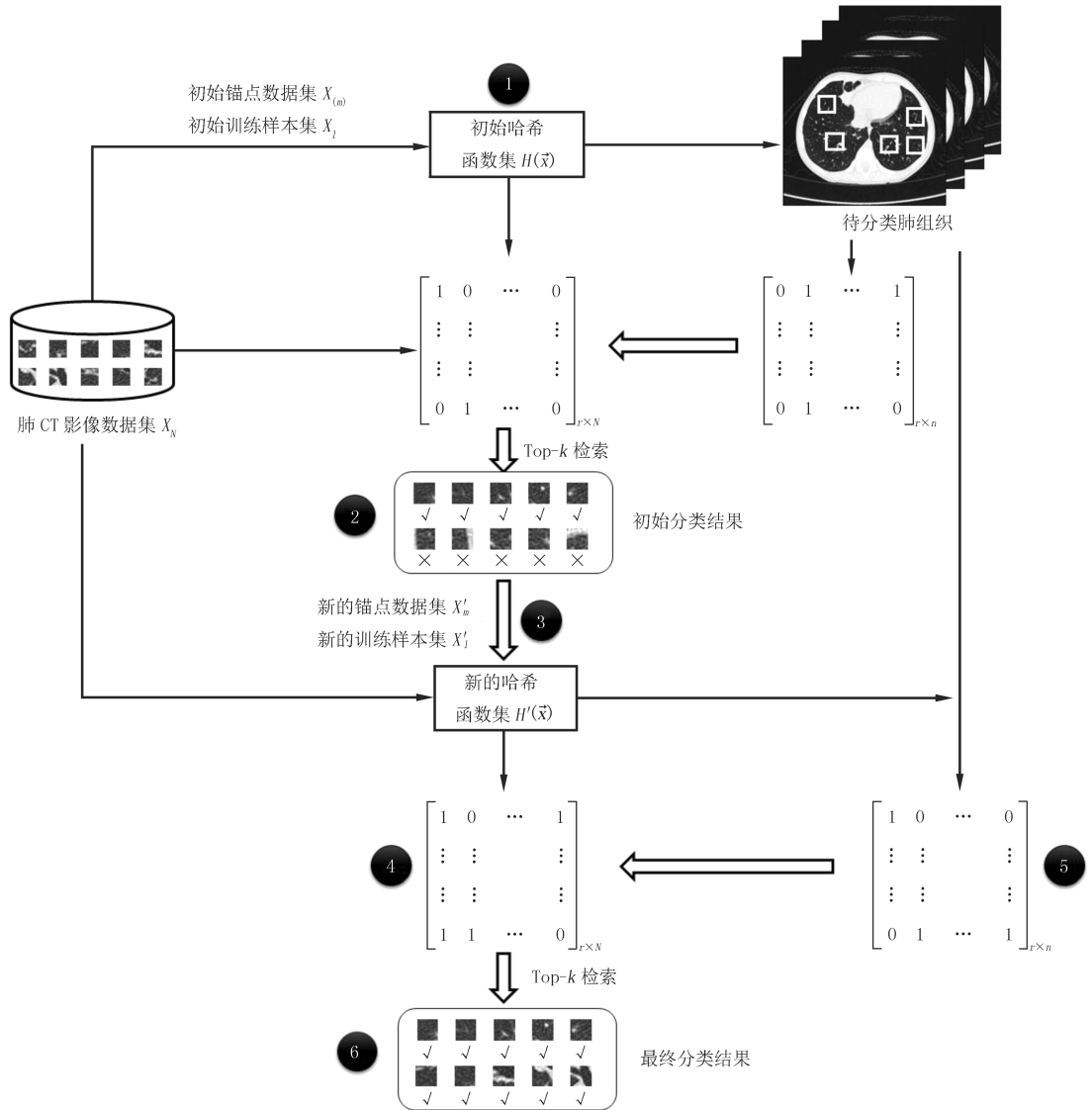


图 2 RFKSH 算法框架

Fig.2 Framework of RFKSH algorithm

### 3 实验

#### 3.1 实验环境与数据集

本文使用 Matlab 2014b 实现上述算法;使用主频为 2.4 GHz 的 Intel Core i5 处理器,内存为 8 GB 的台

式计算机进行实验.

从澳大利亚 Telethon Kids Institute 获取由放射科医生标注的 236 例 CF 患者的肺部 CT 影像;以此为基础,构建 CF 患者肺部组织数据集  $I$ .该数据集包含 16 446 个样例,所有样例的大小均为  $17 \times 17$  像素.表 2 列出了该数据集的统计信息.其中,类别 1、类别 2、类别 3、类别 4、类别 5 和类别 6 包含的样本个数分别为 3 231、3 059、3 383、2 598、1 962 和 2 213 个.

表 2 肺部组织数据集统计  
Tab.2 Statistics on lung tissue dataset

类别	样本数量	类别	样本数量	类别	样本数量
类别 1	3 231	类别 3	3 383	类别 5	1 962
类别 2	3 059	类别 4	2 598	类别 6	2 213

实验中,从中选取 11 512 个样本构成训练集  $X_l$ ,剩余的 4 934 个样本作为测试集  $\bar{X}_l$ .同时,随机从  $X_N$  中选取 500 个样本构成锚点集  $X_m$ .

### 3.2 特征提取

有效的特征表达是进行医学数据分析的关键<sup>[20-22]</sup>.为精确提取肺组织的特征,本文采用如下 4 个步骤所示的特征提取方法.

**步骤 1** 高斯滤波.对肺组织数据集  $I$  中的每一幅图像  $I_i$ ,使用 3 个标准方差分别为  $\delta_1=0.43, \delta_2=0.86, \delta_3=1.30$  的高斯核进行高斯滤波,得到 3 个矩阵  $D_1, D_2, D_3$ .

**步骤 2** 特征抽取.使用如表 3 所示的特征描述子,对上述 3 个矩阵分别进行特征提取.在此基础上,得到特征  $Feat_1, Feat_2$  和  $Feat_3$ .

表 3 特征描述子  
Tab.3 Feature descriptors

特征描述子	含义
$Feat_1$	Hessian 矩阵中每一个像素对应的梯度幅值
$Feat_2$	Hessian 矩阵的特征值
$Feat_3$	Laplacian 算子
$Feat_4$	每一个像素的 HU 值
$Feat_5$	所有 HU 值的均值

同时,对于每一幅图像,将其中每一个像素的 Hounsfield Unit (HU)值和这些 HU 值的平均值分别作为特征  $Feat_4$  和  $Feat_5$ .

**步骤 3** 特征拼接.将步骤 2 中得到的特征向量  $Feat_1, Feat_2, Feat_3, Feat_4$  和  $Feat_5$  拼接成一个长特征向量.

**步骤 4** 数据降维.使用主成分分析(Principal Component Analysis, PCA)法对步骤 3 得到的长特征向量进行降维,得到一个维数为 256 的特征向量  $\bar{x}_i$ .

基于上述步骤,对数据集  $I$  中每一幅图像求得特征向量并得到  $I$  的特征集合  $X_N$ .

### 3.3 实验结果

如第 2 节所述,本文使用 TOP- $k$  方法对肺组织分类:对于测试集  $\bar{X}_l$  中的每一个待分类样例  $X_i$ ,在数据集  $X$  中进行基于内容的图像检索.将检索结果按照与  $X_i$  的相似度由近及远的顺序排列.对于该排列中前  $k$  个检索结果,统计它们对应的类别个数.以此为基础,将  $X_i$  分类为包含检索结果最多的那个类别.

使用平均 AP 值(Mean Average Precision, MAP)作为分类精度指标,评价肺组织分类算法的性能.同时,进行分类实验,比较本文提出的 RFKSH 算法和其他经典的哈希算法 SH, KSH 和 LSH 的性能.

实验中,取 RFKSH 算法的哈希位数  $r$  为 64、 $k$  值为 40;同时,对其他 3 种方法进行优化,使其实验结果取得最优值.表 4 给出了 4 种不同方法对 6 类肺组织进行分类时的 MAP 值.显然,在 6 类肺组织上进行的分类实验,本文提出的 RFKSH 算法都取得了最好的分类结果.

表 4 不同算法的 MAP 值( $r=64, k=40$ )Tab.4 MAP values of different algorithms( $r=64, k=40$ )

	类别 1	类别 2	类别 3	类别 4	类别 5	类别 6	平均精度
RFKSH	0.969 9	0.647 3	0.745 9	0.752 5	0.551 9	0.824 4	0.748 7
KSH	0.828 5	0.526 9	0.511 2	0.492 1	0.385 6	0.628 7	0.562 2
LSH	0.674 8	0.454 5	0.462 2	0.411 3	0.271 9	0.511 2	0.464 3
SH	0.552 3	0.386 5	0.377 2	0.321 1	0.216 6	0.423 6	0.379 6

### 3.4 小 结

由于无监督哈希方法 LSH 和 SH 仅保留原始高维特征的判别信息,没有考虑待分类样例的高层语义特征,因此其分类精确度不高.监督核哈希算法 KSH 引入了核函数和高层语义信息,增强了对线性不可分数据的区分能力,其分类结果好于 LSH 和 SH.本文在 KSH 算法的基础上引入了相关反馈信息,强化了高层语义信息在哈希编码中的特征表达,从而更进一步提高了肺组织的分类精度.

## 4 结 论

本文提出了一种基于相关反馈的监督核哈希方法 RFKSH,用于 CF 患者肺部组织的分类.该方法引入相关反馈机制,在构建哈希编码时,有效组合了肺组织的底层视觉特征和高级语义信息.实验结果表明,同其他经典的哈希算法相比,RFKSH 显著提高了 CF 患者肺部组织的分类精度.

### 参 考 文 献

- [1] Levy H, Kalish L A, Huntington I, et al. Inflammatory markers of lung disease in adult patients with cystic fibrosis[J]. *Pediatric Pulmonology*, 2007, 42(3): 256-262.
- [2] Ciompi F, Palaouroutas A, Loeve M, et al. Lung tissue classification in severe advanced cystic fibrosis from CT scans[C]//Proceedings of the Fourth International Workshop on Pulmonary Image Analysis. Toronto: CreateSpace Independent Publishing Platform, 2011: 57-68.
- [3] He J, Liu W, Chang S, et al. Scalable similarity search with optimized kernel hashing[C]//Proceedings of the ACM Conference on Knowledge Discovery and Data Mining. Washington: ACM Press, 2010: 1129-1138.
- [4] Wang J, Kumar S, Chang S, et al. Semi-supervised hashing for large scale search[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(12): 2393-2406.
- [5] Zhang X, Liu W, Dundar M, et al. Towards large-scale histopathological image analysis: hashing based image retrieval[J]. *IEEE Transactions on Medical Imaging*, 2015, 34(2): 496-506.
- [6] 曾宪华, 袁知洪, 王国胤, 等. 基于多特征多核哈希学习的大规模图像检索[J]. *中国科学: 信息科学*, 2017, 47(1): 1109-1126.
- [7] 柯胜财, 赵永威, 李弼程, 等. 基于卷积神经网络和监督核哈希的图像检索方法[J]. *电子学报*, 2017, 45(1): 157-163.
- [8] Weiss Y, Torralba A, Fergus R. Spectral hashing[C]//Proceedings of Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2008: 1753-1760.
- [9] Kulis B, Grauman K. Kernelized locality-sensitive hashing[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(6): 1092-1104.
- [10] Gong Y, Lazebnik S, Gordo A, et al. Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(12): 2916-2929.
- [11] Liu W, Wang J, Ji R, et al. Supervised hashing with kernels[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE Press, 2012: 2074-2081.
- [12] Liu J, Zhang S, Liu W, et al. Scalable mammogram retrieve using composite anchor graph hashing with iterative quantization[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, 27(1): 2450-2460.
- [13] Zhang X, Liu W, Dundar M, et al. Towards large-scale histopathological image analysis: hashing-based image retrieval[J]. *IEEE Transactions on Medical Imaging*, 2015, 34(2): 496-506.
- [14] Norouzi M, Punjani A, Fleet D J. Fast search in hamming space with multi-index hashing[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE Computer Society Press, 2012: 3108-3115.
- [15] Rui Y, Huang T, Omega M, et al. Relevance feedback: A power tool in interactive content-based image retrieval[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 1998, 8(5): 644-655.
- [16] Shen H L, Tao D C, Ma D F. Dual-force ISOMAP: a new relevance feedback method for medical image retrieval[J]. *PLOS ONE*, 2013,

8(12):1371-1381.

- [17] Huang Y G, Huang H Y, Zhang J. A noisy-smoothing relevance feedback method for content-based medical image retrieval[J]. *Multimedia tools and applications*, 2014, 73(3): 1963-1981.
- [18] Behnam M, Pourghassem H. Optimal query-based relevance feedback in medical image retrieval using score fusion-based classification[J]. *Journal of Digital Imaging*, 2015, 28(2): 160-178.
- [19] Banerjee I, Kurtz C, Devorah A, et al. Relevance feedback for enhancing content-based image retrieval and automatic prediction of semantic image features: application to bone tumor radiographs[J]. *Journal of Biomedical Informatics*, 2018, 84: 123-135.
- [20] 刘艳, 程璐, 孙林. 基于 K-S 检验和邻域粗糙集的特征选择方法[J]. *河南师范大学学报(自然科学版)*, 2019, 47(2): 21-28.
- [21] 徐久成, 黄方舟, 穆辉宇, 等. 基于 PCA 和信息增益的肿瘤特征基因选择方法[J]. *河南师范大学学报(自然科学版)*, 2018, 46(2): 104-110.
- [22] 徐久成, 冯森, 穆辉宇. 基于信噪比与随机森林的肿瘤特征基因选择[J]. *河南师范大学学报(自然科学版)*, 2017, 45(2): 87-92.

## Classification of lung tissue from patients with Cystic Fibrosis disease: A new kernel-based supervised hashing method with relevance feedback

Shen Hualei, Qiu Peng

(College of Computer & Information Engineering, Henan Normal University, Xinxiang 453007, China)

**Abstract:** Performance of traditional hashing methods applied to classification of lung tissues from patients with Cystic Fibrosis (CF) disease is not satisfied, because these methods do not explore discriminant information contained in positive and negative samples obtained from classification results. In this paper, we propose a new kernel-based supervised hashing method boosted by relevance feedback. Firstly, a preliminary hashing function is learned by performing kernel-based supervised hashing on lung tissues. Secondly, the learned hashing function is utilized to encode and classify lung tissues. Then positive and negative samples are obtained from the classification results. Thirdly, a new version of hashing function is learned from these feedback examples. Finally, the newly learned hashing function is employed to classify lung tissues again. Experimental results show that, compared with other existing hashing methods, the proposed method greatly enhances classification results of lung tissue from patients with CF.

**Keywords:** relevance feedback; kernel-based supervised hashing; lung tissue classification; CT image

[责任编辑 陈留院 赵晓华]