



(12) 发明专利申请

(10) 申请公布号 CN 103049526 A

(43) 申请公布日 2013. 04. 17

(21) 申请号 201210559081. X

(22) 申请日 2012. 12. 20

(71) 申请人 中国科学院自动化研究所

地址 100190 北京市海淀区中关村东路 95 号

(72) 发明人 王亮 谭铁牛 赫然 王开业 王威

(74) 专利代理机构 中科专利商标代理有限责任公司 11021

代理人 宋焰琴

(51) Int. Cl.

G06F 17/30 (2006. 01)

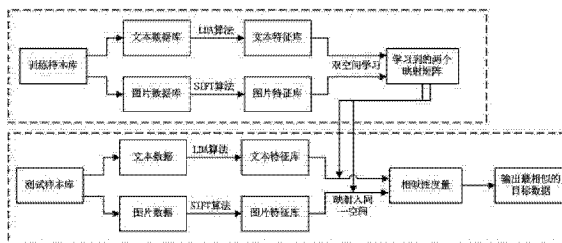
权利要求书 2 页 说明书 4 页 附图 1 页

(54) 发明名称

基于双空间学习的跨媒体检索方法

(57) 摘要

本发明公开了一种基于双空间学习的跨媒体检索方法。该方法包括：首先提取不同模态的多媒体数据的特征，利用双空间学习方法学习得到两个映射矩阵，将不同模态的数据映射到同一个空间；对于测试样本集，将其分为两部分：查询数据集和目标数据集，使用学习得到的映射矩阵将两个数据集的数据映射到统一空间，然后度量查询数据和目标数据之间的距离，并得到与查询数据距离最近的目标数据。本发明可以将不同模态的多媒体数据映射入统一空间进行度量，且在映射的同时进行了特征选择，提高了检索的鲁棒性和准确性，具有良好的运用前景。



1. 一种基于双空间学习的跨媒体检索方法,其特征在于,该方法包括以下步骤:

步骤 S1,收集不同模态的多媒体数据样本,建立跨媒体检索知识数据库,并将所述数据库分为训练集和测试集;

步骤 S2,提取所述数据库中不同模态多媒体数据样本的特征向量;

步骤 S3,基于所述训练集中的不同模态数据的特征向量得到与所述不同模态分别对应的映射矩阵;

步骤 S4,根据所述测试集中多媒体数据样本的模态类别,利用所述步骤 S3 得到的映射矩阵将它们映射到同一空间;

步骤 S5,将映射到同一空间后的测试集中同一模态类别的多媒体数据作为查询集,另一模态类别的多媒体数据作为目标集;

步骤 S6,对于所述查询集中的一个多媒体数据,根据多媒体数据之间的相似度,得到所述目标集中与它最为相似的多媒体数据,从而得到跨媒体检索结果。

2. 根据权利要求 1 所述的方法,其特征在于,所述不同模态为双模态。

3. 根据权利要求 2 所述的方法,其特征在于,所述双模态为文本和图像。

4. 根据权利要求 3 所述的方法,其特征在于,对于文本多媒体数据样本,提取其隐狄雷克雷分布特征向量;对于图像多媒体数据样本,提取其尺度不变特征变换特征向量。

5. 根据权利要求 1 所述的方法,其特征在于,所述步骤 S3 进一步包括以下步骤:

步骤 S31,基于所述训练集中不同模态数据的特征向量建立目标函数;

步骤 S32,求解所述目标函数得到与所述不同模态分别对应的映射矩阵。

6. 根据权利要求 5 所述的方法,其特征在于,所述目标函数为:

$$\min_{\mathbf{U}_a, \mathbf{U}_b} \frac{1}{2} (\|\mathbf{X}_a^T \mathbf{U}_a - \mathbf{Y}\|_F^2 + \|\mathbf{X}_b^T \mathbf{U}_b - \mathbf{Y}\|_F^2) + \lambda_1 (\|\mathbf{U}_a\|_{21} + \|\mathbf{U}_b\|_{21}) + \lambda_2 \|\mathbf{X}_a^T \mathbf{U}_a - \mathbf{X}_b^T \mathbf{U}_b\|_*,$$

其中, \mathbf{U}_a 和 \mathbf{U}_b 为与双模态分别对应的两个映射矩阵, λ_1 和 λ_2 均为待定参数, $\|\cdot\|_F$ 为求取 F-范数操作, $\|\cdot\|_{21}$ 为求取行向量二范数和操作, $\|\cdot\|_*$ 为求取迹范数操作。

7. 根据权利要求 5 所述的方法,其特征在于,对于所述目标函数的求解使用迭代算法来进行。

8. 根据权利要求 7 所述的方法,其特征在于,所述步骤 S32 进一步包括以下步骤:

步骤 S321,设置映射矩阵 \mathbf{U}_a 和 \mathbf{U}_b 的初始值为零矩阵,设置最大迭代次数 k,迭代次数的初始值为 1;

步骤 S322,对矩阵 $(\mathbf{X}_a^T \mathbf{U}_a \mathbf{U}_a^T \mathbf{X}_a + \mathbf{X}_b^T \mathbf{U}_b \mathbf{U}_b^T \mathbf{X}_b)$ 进行特征值分解,得到特征值分解结果: $\mathbf{V} \text{Diag}(d) \mathbf{V}^T$, 其中, $\mathbf{X}_a = [\mathbf{x}_1^a, \mathbf{x}_2^a, \dots, \mathbf{x}_n^a] \in \mathbb{R}^{d1 \times n}$ 为所述训练集中文本多媒体数据的特征向量组成的矩阵, d1 为特征向量的维数, n 为特征向量的个数, a 表示模态类别; $\mathbf{X}_b = [\mathbf{x}_1^b, \mathbf{x}_2^b, \dots, \mathbf{x}_n^b] \in \mathbb{R}^{d2 \times n}$ 为所述训练集中图像多媒体数据的特征向量组成的矩阵, d2 为特征向量的维数, n 为特征向量的个数, b 表示模态类别; \mathbf{V} 为以矩阵 $(\mathbf{X}_a^T \mathbf{U}_a \mathbf{U}_a^T \mathbf{X}_a + \mathbf{X}_b^T \mathbf{U}_b \mathbf{U}_b^T \mathbf{X}_b)$ 的特征值向量为列向量组成的矩阵; d 为特征值向量对应的特征值组成的向量;

步骤 S323,设定中间变量 $\mathbf{S}^{-1} = \mathbf{V} \text{Diag}(c) \mathbf{V}^T$, 其中, c 是一个向量,它的第 i 个元素为 $c_i = 1 / \sqrt{d_i + \mu}$, 其中, d_i 是向量 d 的第 i 个元素, μ 为一极小数;

步骤 S324,根据下式计算中间向量 p 和 q 的第 i 个元素 p_i 和 q_i :

$$\begin{cases} p_i = \frac{1}{2\sqrt{\|\mathbf{u}_a^i\|_2^2} + \varepsilon}, \\ q_i = \frac{1}{2\sqrt{\|\mathbf{u}_b^i\|_2^2} + \varepsilon} \end{cases},$$

并令 $\mathbf{P} = \text{Diag}(p)$, $\mathbf{Q} = \text{Diag}(q)$, 其中, ε 为一极小数;

步骤 S325, 通过求解下式中的两个线性问题来求得所述映射矩阵 \mathbf{U}_a 和 \mathbf{U}_b :

$$\begin{cases} (\mathbf{X}_a \mathbf{X}_a^T + \lambda_1 \mathbf{P}_a + \lambda_2 \mathbf{X}_a \mathbf{S}^{-1} \mathbf{X}_a^T) \mathbf{U}_a = \mathbf{X}_a \mathbf{Y} \\ (\mathbf{X}_b \mathbf{X}_b^T + \lambda_1 \mathbf{P}_b + \lambda_2 \mathbf{X}_b \mathbf{S}^{-1} \mathbf{X}_b^T) \mathbf{U}_b = \mathbf{X}_b \mathbf{Y} \end{cases},$$

其中, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n] \in \mathbb{R}^{n \times c}$ 为类别标签矩阵, n 为特征向量的个数, c 为模态类别的个数;

步骤 S326, 判断此时的迭代次数是否小于最大迭代次数 k , 若是, 则转向所述步骤 S322 继续迭代; 若否, 则停止迭代, 得到所述映射矩阵 \mathbf{U}_a 和 \mathbf{U}_b 。

9. 根据权利要求 1 所述的方法, 其特征在于, 所述多媒体数据之间的相似度使用欧式距离来度量。

基于双空间学习的跨媒体检索方法

技术领域

[0001] 本发明涉及模式识别技术领域,特别是一种基于双空间学习的跨媒体检索方法。

背景技术

[0002] 近些年来,随着 Web2.0 技术的兴起,互联网上的多媒体数据呈现出爆炸式的增长,用户需要从这些海量数据中检索自己想要的文章、图片、视频等内容。而现在的主流搜索引擎如百度、谷歌等,采用的搜索技术依然是基于关键字的,而这种搜索技术无法解决各个媒体之间存在的语义鸿沟。为了实现跨媒体检索,需要提供一种从一类多媒体数据检索另一类多媒体数据的综合检索方法。

[0003] 跨媒体检索方法的研究已经引起了国内外学术界的重视。目前国际上已经提出了一些方法来建模不同模态的多媒体数据之间的关系,现有方法可以分为两类:1) 将两个模态的多媒体数据映射到同一子空间,然后在这个学习到的子空间里度量不同模态的多媒体数据之间的相似性,以此来实现跨媒体检索。2) 将一类模态的多媒体数据转换到另一种模态的多媒体数据的特征空间中进行不同模态数据的相似性度量,以此来实现跨媒体检索。然而,现有方法都忽略了另一个重要问题,即如何从不同模态的特征空间中同时选择出最相关和最有区分力的特征,这个问题称为双空间特征选择。为了在学习子空间的同时,进行双空间特征选择,本发明提出一种基于双空间学习的跨媒体检索方法,该方法有效地实现了不同模态媒体之间的语义鸿沟的跨越,进而使得搜索引擎返回的结果更加准确。

发明内容

[0004] 为了解决现有技术存在的问题,本发明的目的是提供一种基于双空间学习的跨媒体检索方法。该方法通过最小化一个目标函数,能同时进行子空间学习和双空间特征选择,在学习到的子空间中度量不同模态的多媒体数据的相似性,以此来实现跨媒体检索。

[0005] 本发明提出的一种基于双空间学习的跨媒体检索方法包括以下步骤:

[0006] 步骤 S1,收集不同模态的多媒体数据样本,建立跨媒体检索知识数据库,并将所述数据库分为训练集和测试集;

[0007] 步骤 S2,提取所述数据库中不同模态多媒体数据样本的特征向量;

[0008] 步骤 S3,基于所述训练集中的不同模态数据的特征向量得到与所述不同模态分别对应的映射矩阵;

[0009] 步骤 S4,根据所述测试集中多媒体数据样本的模态类别,利用所述步骤 S3 得到的映射矩阵将它们映射到同一空间;

[0010] 步骤 S5,将映射到同一空间后的测试集中同一模态类别的多媒体数据作为查询集,另一模态类别的多媒体数据作为目标集;

[0011] 步骤 S6,对于所述查询集中的一个多媒体数据,根据多媒体数据之间的相似度,得到所述目标集中与它最为相似的多媒体数据,从而得到跨媒体检索结果。

[0012] 根据本发明的方法,可以将不同模态的多媒体数据映射入统一空间进行度量,且

在映射的同时进行了特征选择,提高了检索的鲁棒性和准确性,具有良好的运用前景。本发明方法能有效地跨越了不同模态媒体之间的语义鸿沟,进而使得跨媒体搜索引擎返回的结果更加准确。

附图说明

[0013] 图 1 是本发明基于双空间学习的跨媒体检索方法流程图;

[0014] 图 2 是根据本发明跨媒体检索方法从文本到图像的检索效果示意图。

具体实施方式

[0015] 为使本发明的目的、技术方案和优点更加清楚明白,以下结合具体实施例,并参照附图,对本发明进一步详细说明。

[0016] 本发明通过双空间学习来学习两个映射矩阵,将不同模态的数据映射到同一个子空间中去,并在映射的同时进行了特征选择;在学习到的子空间中度量不同模态的数据的相似性,以此来实现跨媒体检索的目的。

[0017] 图 1 是本发明基于双空间学习的跨媒体检索方法流程图,如图 1 所示,本发明提出的一种基于双空间学习的跨模态检索方法包括以下几个步骤:

[0018] 步骤 S1,收集不同模态的多媒体数据样本,建立跨媒体检索知识数据库,并将所述数据库分为训练集和测试集;

[0019] 所述不同模态比如可以为文本、图像等模态,为了便于对本发明进行说明,下文以文本和图像这两个模态的数据为例进行解释。

[0020] 在将所述数据库分为训练集和测试集时,可根据实际需要进行划分,比如可将所述数据库中的 80% 的数据划分为训练集,剩下的 20% 的数据划分为测试集。

[0021] 步骤 S2,提取所述数据库中不同模态多媒体数据样本的特征向量;

[0022] 本发明中,对于文本和图像分别使用隐狄克雷分布 (Latent Dirichlet Allocation, LDA) 算法和尺度不变特征变换 (Scale-Invariant Feature Transform, SIFT) 算法进行特征提取。

[0023] 步骤 S3,基于所述训练集中的不同模态数据的特征向量得到与所述不同模态分别对应的映射矩阵;

[0024] 在对该步骤进行详细介绍前,先介绍一些矩阵运算符:

[0025] 对于矩阵 $\mathbf{M} \in \mathbb{R}^{n \times m}$, $\mathbb{R}^{n \times m}$ 为 $n \times m$ 的矩阵空间,矩阵 \mathbf{M} 的第 i 行元素表示为 \mathbf{m}^i ,矩阵 \mathbf{M} 的第 j 列元素表示为 \mathbf{m}_j ,而矩阵 \mathbf{M} 的 Frobenius 范数,简称 F-范数定义为:

$$[0026] \quad \|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^n \|\mathbf{m}^i\|_2^2} \quad (1)$$

[0027] 矩阵 \mathbf{M} 的行向量的二范数的和 $\|\mathbf{M}\|_{21}$ 定义为:

$$[0028] \quad \|\mathbf{M}\|_{21} = \sum_{i=1}^n \|\mathbf{m}^i\|_2 \quad (2)$$

[0029] 矩阵 \mathbf{M} 的迹范数定义为:

$$[0030] \quad \|\mathbf{M}\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i \quad (3)$$

[0031] 其中, σ_i 表示矩阵 M 的第 i 个奇异值。

[0032] 对于向量 $\mathbf{u} \in \mathbb{R}^m$, $\text{Diag}(\mathbf{u}) \in \mathbb{R}^{m \times m}$ 表示一个对角线元素为 u_i 的对角矩阵, 其中, u_i 表示向量 \mathbf{u} 的第 i 个元素。

[0033] 设所述训练集中文本多媒体数据的特征向量组成的矩阵为 $\mathbf{X}_a = [\mathbf{x}_1^a, \mathbf{x}_2^a, \dots, \mathbf{x}_n^a] \in \mathbb{R}^{d1 \times n}$, $d1$ 为特征向量的维数, n 为特征向量的个数, a 表示模态类别; 图像多媒体数据的特征向量组成的矩阵为 $\mathbf{X}_b = [\mathbf{x}_1^b, \mathbf{x}_2^b, \dots, \mathbf{x}_n^b] \in \mathbb{R}^{d2 \times n}$, $d2$ 为特征向量的维数, n 为特征向量的个数, b 表示模态类别; 类别标签矩阵为 $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n] \in \mathbb{R}^{n \times c}$, n 为特征向量的个数, c 为模态类别的个数。

[0034] 所述步骤 S3 进一步包括以下步骤:

[0035] 步骤 S31, 基于所述训练集中不同模态数据的特征向量建立目标函数:

$$[0036] \quad \min_{\mathbf{U}_a, \mathbf{U}_b} \frac{1}{2} (\|\mathbf{X}_a^T \mathbf{U}_a - \mathbf{Y}\|_F^2 + \|\mathbf{X}_b^T \mathbf{U}_b - \mathbf{Y}\|_F^2) + \lambda_1 (\|\mathbf{U}_a\|_{21} + \|\mathbf{U}_b\|_{21}) + \lambda_2 \|\mathbf{X}_a^T \mathbf{U}_a - \mathbf{X}_b^T \mathbf{U}_b\|_* \quad (4)$$

[0037] 其中, \mathbf{U}_a 和 \mathbf{U}_b 为与文本和图像分别对应的两个映射矩阵, λ_1 和 λ_2 均为待定参数, 取值范围为 0.0001 至 0.1;

[0038] 步骤 S32, 求解所述目标函数 (比如通过迭代算法) 得到与所述不同模态分别对应的映射矩阵:

[0039] 所述步骤 S32 进一步包括以下步骤:

[0040] 步骤 S321, 设置映射矩阵 \mathbf{U}_a 和 \mathbf{U}_b 的初始值为零矩阵, 设置最大迭代次数 k , 并设置迭代次数的初始值为 1;

[0041] 步骤 S322, 对矩阵 $(\mathbf{X}_a^T \mathbf{U}_a \mathbf{U}_a^T \mathbf{X}_a + \mathbf{X}_b^T \mathbf{U}_b \mathbf{U}_b^T \mathbf{X}_b)$ 进行特征值分解, 得到特征值分解结果: $\mathbf{V} \text{Diag}(\mathbf{d}) \mathbf{V}^T$, 其中, \mathbf{V} 为以矩阵 $(\mathbf{X}_a^T \mathbf{U}_a \mathbf{U}_a^T \mathbf{X}_a + \mathbf{X}_b^T \mathbf{U}_b \mathbf{U}_b^T \mathbf{X}_b)$ 的特征值向量为列向量组成的矩阵; \mathbf{d} 为特征值向量对应的特征值组成的向量。

[0042] 步骤 S323, 设定中间变量 $\mathbf{S}^{-1} = \mathbf{V} \text{Diag}(\mathbf{c}) \mathbf{V}^T$, 其中, \mathbf{c} 是一个向量, 它的第 i 个元素为 $c_i = 1/\sqrt{d_i + \mu}$, 其中, d_i 是向量 \mathbf{d} 的第 i 个元素, μ 为一极小数。

[0043] 步骤 S324, 根据下式计算中间向量 \mathbf{p} 和 \mathbf{q} 的第 i 个元素 p_i 和 q_i :

$$[0044] \quad \begin{cases} p_i = \frac{1}{2\sqrt{\|\mathbf{u}_a^i\|_2^2 + \varepsilon}} \\ q_i = \frac{1}{2\sqrt{\|\mathbf{u}_b^i\|_2^2 + \varepsilon}} \end{cases} \quad (5)$$

[0045] 并令 $\mathbf{P} = \text{Diag}(\mathbf{p})$, $\mathbf{Q} = \text{Diag}(\mathbf{q})$, 其中, ε 为一极小数。

[0046] 步骤 S325, 通过求解下式中的两个线性问题来求得所述映射矩阵 \mathbf{U}_a 和 \mathbf{U}_b :

$$[0047] \quad \begin{cases} (\mathbf{X}_a \mathbf{X}_a^T + \lambda_1 \mathbf{P}_a + \lambda_2 \mathbf{X}_a \mathbf{S}^{-1} \mathbf{X}_a^T) \mathbf{U}_a = \mathbf{X}_a \mathbf{Y} \\ (\mathbf{X}_b \mathbf{X}_b^T + \lambda_1 \mathbf{P}_b + \lambda_2 \mathbf{X}_b \mathbf{S}^{-1} \mathbf{X}_b^T) \mathbf{U}_b = \mathbf{X}_b \mathbf{Y} \end{cases} \quad (6)$$

[0048] 步骤 S326, 判断此时的迭代次数是否小于最大迭代次数 k , 若是, 则转向所述步骤 S322 继续迭代; 若否, 则停止迭代, 得到所述映射矩阵 \mathbf{U}_a 和 \mathbf{U}_b 。

[0049] 步骤 S4, 根据所述测试集中多媒体数据样本的模态类别, 利用所述步骤 S3 得到的映射矩阵将它们映射到同一空间;

[0050] 步骤 S5, 将映射到同一空间后的测试集中同一模态类别的多媒体数据作为查询

集,另一模态类别的多媒体数据作为目标集;

[0051] 步骤 S6,对于所述查询集中的一个多媒体数据,根据多媒体数据之间的相似度,得到所述目标集中与它最为相似的多媒体数据,从而得到跨媒体检索结果。

[0052] 所述多媒体数据之间的相似度可以使用欧式距离来度量,对于欧式距离来说,距离越近就说明这两个多媒体数据越相似。

[0053] 为了验证本发明方法的有效性,接下来将本发明所提出的方法应用于一个多媒体数据库。该数据库中包含 10 个语义范畴的多媒体数据,包括 2866 对相同语义信息的文本和图像,随机选取其中的 2173 对文本和图像样本作为训练集,选取其中的 693 对文本和图像样本作为测试集,应用本发明方法的具体步骤如下:

[0054] 1) 对于所述数据库中的数据根据模态类别的不同提取不同的特征向量:对于文本模态,提取 10 维的 LDA 语义特征,对于图像模态,提取 128 维的 SIFT 特征;

[0055] 2) 将所述训练集中文本数据的特征向量和图像数据的特征向量组成特征矩阵,构造目标函数,通过最小化目标函数来学习得到与文本和图像对应的两个映射矩阵,所述映射矩阵的学习过程是通过一个迭代过程实现的;

[0056] 3) 根据所述模态类别的不同,将测试集中的文本数据的特征向量和图像数据的特征向量通过学习得到的映射矩阵映射到同一空间中;

[0057] 4) 将测试集中的文本数据作为查询集,图像数据作为目标集,给定一个文本文档,能够得到目标集中和它最相似(即欧式距离最近)的图像。图 2 是根据本发明跨媒体检索方法从文本到图像的检索效果示意图,图 2 左侧是两个文本文档,右侧是根据本发明方法得到的与这两个文本文档最为相似的图像。

[0058] 综上,本发明提出了一种有效的基于双空间学习的跨媒体检索方法。本发明易于实现、性能稳定,识别准确率高。本发明提出有效地跨越了不同模态媒体之间的语义鸿沟,与传统的跨媒体方法比较,本发明表现出更广泛的运用前景和更大的市场价值。

[0059] 以上所述的具体实施例,对本发明的目的、技术方案和有益效果进行了进一步详细说明,所应理解的是,以上所述仅为本发明的具体实施例而已,并不用于限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

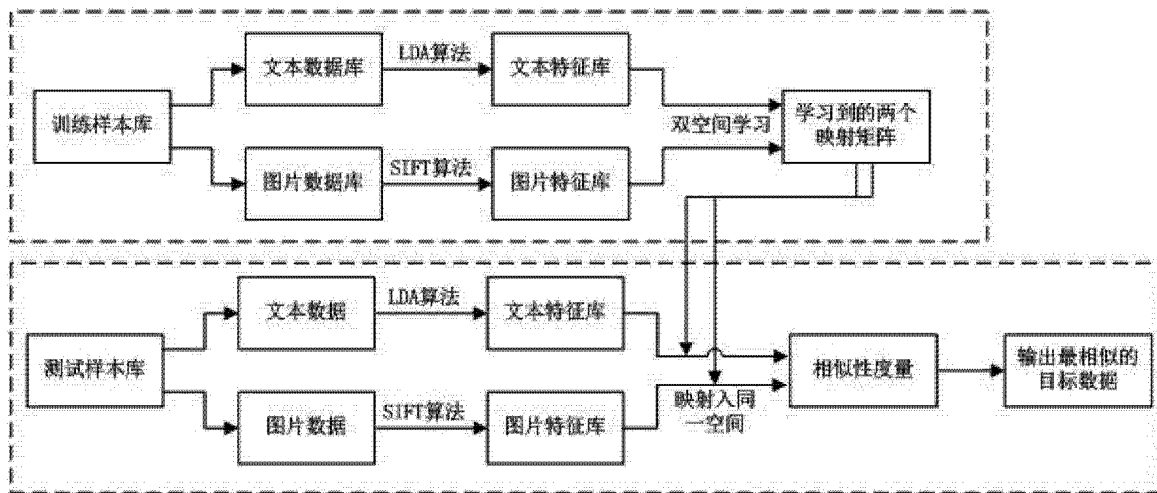
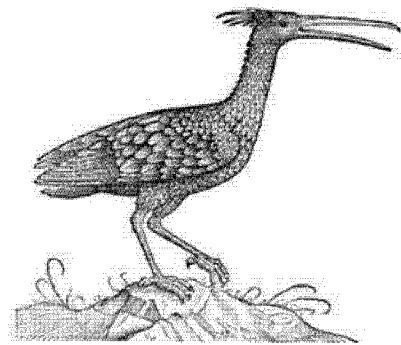


图 1

The Kakapo is the only species of flightless parrot in the world, and the only flightless bird that has a lek breeding system. "Collins Field Guide to New Zealand Wildlife", Terrence Lindsey and Rod Morris, Harper Collins Publishers (New Zealand) limited, 2000. Males loosely gather in an arena and compete with each other to attract females. Females watch the males display, or "lek". Merton, D.V. (1976). Conservation of the kakapo: a progress report. In "Proc. Science in Nat. Parks". National Parks Authority, Wellington, N.Z. National Parks Series No. 6: 139-148. They choose a mate based on the quality of his display; they are not pursued by the males in any overt way. No pair bond is formed; males and females meet only to mate. During the courting season, males leave their home ranges for hilltops and ridges where they establish their own mating courts. These leks can be up to 7 kilometres (4 mi) from a Kakapo's usual territory and are an average of 50 metres (160ft) apart within the lek arena [...].



The Byzantines were unable to respond effectively to the Muslim advance in Africa, because the two decades between 695 and 715 were a period of great domestic turmoil. They did react with raids of their own in the East, such as the one in 709 against Egypt which captured the local admiral, but they also were aware of a coming onslaught: as Caliph al-Walid I (705-715) readied his forces for a renewed assault against Constantinople, Emperor Anastasios II (713-715) prepared the capital, and mounted an unsuccessful preemptive strike against the Muslim naval preparations. Anastasios was soon overthrown by Theodosius III (715-717), who in turn was replaced, just as the Muslim army was advancing through Anatolia, by Leo III the Isaurian (717-741). It was Leo III who faced the second and final Arab siege of Constantinople. The use of Greek fire, which devastated the Arab fleet, was again instrumental in the Byzantine victory, while a harsh winter and Bulgar attacks further sapped the besiegers' strength. [...]

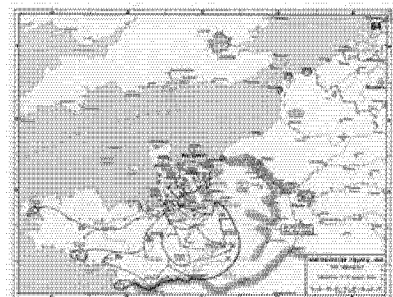


图 2