

基于形式概念分析的图像检索的研究

白津宁

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)
(*通信作者电子邮箱 2624887157@qq.com)

摘要: 为了提高图像检索的效率, 提出了基于形式概念分析的图像检索技术。形式概念分析作为数据分析、规则提取和知识发现的强有力工具, 将数据用概念格的形式有效地结合起来, 从而发现概念内部的隐含知识。以形式化概念分析为理论基础, 给出了图像的形式背景的定义, 生成了图像的概念格, 同时给出了图匹配的相关知识。将图像的形式背景生成概念格之后, 将概念格相对应的 Hasse 图, 然后利用图匹配找到与示例图像相似的图像, 最后利用概念格相似性进行图匹配的计算。即将形式概念分析(主要是概念格)应用于图像检索中进行图像特征相似度的计算。

关键词: 图像检索; 形式概念分析理论; 概念格; 相似性

中图分类号: P301.6 **文献标志码:** A

Research on Image Retrieval Based on Concept Analysis

BAI JinNing

(Institute of Computer Science and Technology, Dalian Maritime University, Dalian Liaoning 116026, China)

Abstract : In order to improve the efficiency of image retrieval, the paper proposes a technique of image retrieval based on formal concept analysis. As a powerful tool for data analysis, rule extraction and knowledge discovery, FCA combines data effectively by using the concept lattice by which we can find the implicit knowledge of the concept inside. Based on the theory formal concept analysis the paper gives the definition of the formal background of the image, generates the concept lattice of the image, and presents the related knowledge of graph matching. After generating the concept lattice based on the formal context of images, the Hasse diagram corresponding with the concept lattice can be generated and the images which are similar to the sample image also be found by lattice matching method. The concept lattice similarity calculation methods are used to graph matching. Recently the formal concept analysis (mainly the concept of grid) has been widely applied to the fields of the machine learning, the data mining and the retrieval information calculation of the similarity of the image features.

Key words: Image Retrieval; Formal Concept Analysis; Concept Lattice; Similarity

0 引言

随着多媒体技术及网络技术的迅速发展, 多媒体数据大量涌入了人们的生活, 彩色图像在我们的生活中占据着越来越重要的地位。人们所熟知的淘宝、京东等电子购物网站, 都是将物品以图片的形式展现给顾客, 需要人们按照自己的要求对符合条件的图片进行查找, 也就是我们通常所说的图像检索。起初的图像检索技术是基于文本的图像检索, 主要是利用文本对图像的特征进行描述, 其主要包

括: 图像的标号、作者、名称信息等^[1]。

后期人们开始基于内容的图像检索, 主要是利用图像底层特征, 如颜色、纹理、形状等特征对图像进行检索。人们在进行图像检索时, 首先要对图像进行处理, 然后建立图像数据库和图像特征数据库, 系统对录入的图像库进行分析, 对图像进行分类, 并且统一建模, 最后把按照不同的图像模型提取的图像特征(如颜色、形状、纹理)放入图像特征库中^[2], 并且对图像特征库建立索引用来提高查询效率。

收稿日期: 2015-10-30; 修回日期: 2015-11-26。

基金项目: 国家自然科学基金资助项目(61370168); 国家青年科学基金资助项目(61402316)。

作者简介: 白津宁(1992 —), 女, 辽宁阜新人, 硕士, 主要研究方向: 概念格、数据挖掘、图像处理。

人们在经过接口输入查询条件时，能够利用一种图像特征或者几种图像特征的组合来表示，然后系统根据特征相似匹配度算法，计算示例图像关键特征与特征库中图像特征的相似度，然后系统将匹配图像反馈给客户。

形式概念分析(Formal Concept Analysis, FCA)模型是一种有效的表示数据和获取知识的工具，已成功地应用于数据分析和知识提取等领域^[3]。以人的认知为中心，提供了一种与传统的、统计的数据分析和知识表示完全不同的方法，成为了人工智能学科的重要研究对象。形式概念分析核也的结构是概念格，概念格是巧来描述概念和属性之间的关系，并说明概念和概念之间泛化和例化的关系的。这样就可利用概念格建立概念结构层次上数据的关系模型，而与概念格相对应的H a s s e图实现了数据的可视化，形式概念分析在机器学习、数据挖掘、信息检索等领域得到了广泛的应用。

本文在形式概念分析理论上，提出了基于形式概念分析的图像检索，利用形式概念分析中概念格相似度的计算，给出了图像相似性匹配的新方法。利用形式概念分析对图像进行了两次检索；第一次检索是对图像库进行分类，这在很大程度上减少了图像检索耗费的时间；第二次检索，是对示例图像和图像库中的每一幅图像进行概念格匹配，这使得图像检索的准确率大大提高。

1 FCA 相关概念

FCA 中两个主要的概念是形式背景和形式概念，在一个抽象的形式背景上，所有的形式概念生成一种概念格结构。本文涉及到的相关概念如下：

一个抽象的形式背景定义为 $K=(O, M, R)$ ，由对象集 O 和属性集 M 以及 O 与之间的关系 R 组成，且 $R \subseteq O \times M$ 。其中， $h=(P, Q)$ 为 K 上的任一个二元组， $P \subseteq O, Q \subseteq M$ ，而且满足映射关系：

$$f(P)=\{m\in M|\forall\ o\in P,\ oRm\}$$
$$g(Q)=\{o\in O|\forall\ m\in Q,\ oRm\}$$

如果 $f(P)=Q, g(Q)=P$ ，则意味着二元组 (P, Q) 满足外延内涵的最大扩展性，那么定义 $h=(P, Q)$ 为 K 上的一个形式概念， Q 被称作概念 h 的内涵， P 被称作概念 h 的外延。

设 $h_1=(P_1, Q_1), h_2=(P_2, Q_2)$ 和 $h_3=(P, Q)$ 是 K 上的三个形式概念，如果 $P_1 \subseteq P_2$ (或 $Q_1 \supseteq Q_2$)，则称 h_2 是 h_1 的祖先概念， h_1 是 h_2 的子孙概念，记为 $h_1 \leq h_2$ ，其中“ \leq ”表示为概念之间的“层次序”。特别地，若不存在形式概念 h_3 ，有 $h_1 \leq h_3 \leq h_2$ 成立，则称概念 h_2 是 h_1 的直接父概念， h_1 是 h_2 的直接子概念。 K 中的所有形式概念，用这种层次(偏)序组成的集合称为 K 上的概念格，标记为 $\langle L(O, M, R), \leq \rangle$ ，为描述方便，简记为 L 。

2 基于 FCA 的图像检索研究

在对一个庞大的图像数据库进行检索前，首先利用形式概念分析和语义特征对检索图像库进行分类（第一次检索），这样可以对图像数据库实现宏分类，减少我们的计算量，即节约了计算机的运行时间；然后对示例图像和示例图像所属的类别建立形式背景和概念格，利用概念格的相似性，计算图像之间的相似度，这样可以使检索结果更精确。概念格是形式概念分析的基础理论，是知识挖掘和数据分析的强有力工具，已经在信息检索、软件工程、知识发现等方面取得了很好的应用效果。建立概念格的过程实际上是概念聚类过程。概念格描述了对对象和特征之间的关系，用于识别数据集中概念结构的数据分析理论。概念格可以用来表示概念之间的关系，并且与概念格相对应的H a s s e图实现了数据的可视化。生成概念格并通过概念格构造H a s s e图的过程，其实就是聚类和分类的过程，而且H a s s e图实现了概念之间关系的可视化。本

文利用形式概念的优势，对图像进行分类和检索，可以提高检索的有效性^[4]。

2. 1 概念格的相似度

每一个概念格都是由概念组成的，概念格上的结点就是概念，要计算两个概念格的相似程度，实际上是在计算概念格上的结点（概念）之间的相似度。概念格上结点（概念）之间的距离在某种程度上反应了概念的相似性，距离小的结点的相似度大于结点距离大的结点之间的相似度。在同一概念格中，概念之间存在继承关系和二元关系^[5]，因此我们可以用继承关系或者二元关系链的长度来度量概念的距离。

两个概念的距离主要由它们自己的深度（概念距离根节点的距离）和共同父概念的深度（共同父概念到根节点的距离）来决定。概念的深度越大，概念表达的越是明确，两个概念距离共同父概念的距离越小，两个概念的关系越密切。具体表达式如下：

$$d(A,B)=\frac{d(A,O)+d(B,O)}{2\times d(LCA((A,B)O))}$$

其中 A,B 为两个概念，O 为根节点，d(A,O) 表示概念 A 距离根节点的距离，d(B,O) 表示概念 B 距离根节点的距离，LCA(A,B) 为概念 A,B 的共同父概念。例 2. 1. 下图是一个概念格，计算概念的距离。

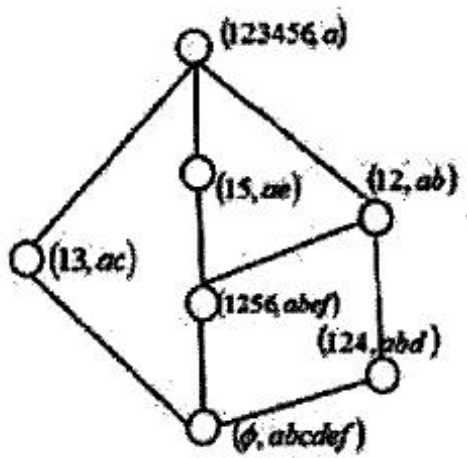


图 2. 1 概念格

对于概念 A=(12, ab) 和概念 B= (15, af) 的距离，先找到 A,B 的共同父概念为 LCA(A,B) = (1256, abef)，它到根节点为 1，概念 A 到根节点的距离为 3，概念 B 到根节点的距离为 3，有 A,B 的距离公式得：

$$d(A,B)=\frac{d(A,O)+d(B,O)}{2\times d(LCA(A,B)O)}=\frac{3+3}{2\times 1}=3$$

2. 2 用 FCA 实现图像第一次检索

在对图像进行检索之前，我们先对图像库进行分类。在检索图像时，先提取图像的语义信息对图像进行粗略的分类。首先根据图像的语义特征建立一个图像数据库。我们对图像库中的图像进行标号，而且把图像作为形式背景中的对象，分别记为 F₁, F₂, F₃, ..., F_N；对图像中出现的事物，例如：动物、人、花草、树木、建筑物等也进行标号，并把这些语义特征作为形式背景中的属性，分别记为 T₁, T₂, T₃, ..., T_M。

例 2. 2 这里本文将 F=(F₁, F₂, F₃, ..., F_N) 作为对象集合，T=(T₁, T₂, T₃, ..., T_M) 为属性集合，对图像库建立形式背景，其中 " 1 " 表示图像中含有某个属性，反之，" 0 " 则表示围像中不含该属性。

表 2. 1 图像库的形式背景

对象 图像	T ₁	T ₂	T ₃	...	T _M
F ₁	0	0	1	1	0
F ₂	0	1	0	0	0
F ₃	0	0	0	0	1
...	0	0	0	0	1
F _N	1	0	1	0	0

利用图像的形式背景，对图像库生成概念格。这里对图像库中所有的图像以及示例图像在同一个形式背景下生成概念格，用概念格的相似性，来计算图像的相似性。从而实现图像库的分类。以人物，大海，岩石，树木等为背景分类如下：

表 2.2 图像的形式背景

	a. 人物	b. 大海	c. 岩石	d. 树木
图像 1	1	0	0	0
图像 2	0	1	0	0
图像 3	1	0	1	0
图像 4	1	1	0	0
图像 5	0	1	0	1
图像 6	1	1	1	0

根据图像的全部形式背景生成概念格，每一个概念格都是一个概念。

表 2.3 图像形式背景生成的全部概念格

C_1	{图像 1; 2; 3; 4; 5; 6}	{}
C_2	{图像 2; 4; 5; 6}	{b}
C_3	{图像 5}	{b, d}
C_4	{图像 1; 3; 4; 6}	{a}
C_5	{图像 3; 6}	{a, c}
C_6	{图像 4; 6}	{a, b}
C_7	{图像 6}	{a, b, c}
C_8	{}	{a, b, c, d}

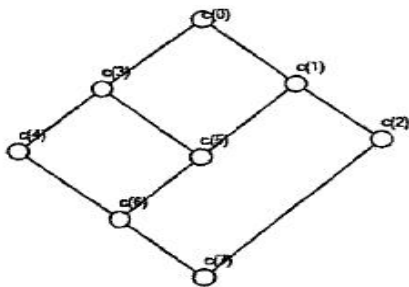


图 2.2 图像形式背景生成的全部概念格

通过以上形式背景的建立和概念格的生成，我们可以很清楚的看到图片（对象）和语义特征属性）之间的关系，因此我们可利用图像的概念格对图像进行第一次检索。

2. 3 用 FCA 实现图像第二次检索

我们利用概念格的相似度对图像进行二次检索。每一幅图像都可以生成一个概念格，要想从图像集中找到与示例图像相似的图像，就是要使图像集中每个图像生成的概念格与示例图像的概念格相匹配。通过计算概念格之间的相似度，实现图像概

念格的匹配问题，然后按照相似度的大小进行排序，就完成了图像的二次检索。

基于形式概念分析的图像检索系统如下：

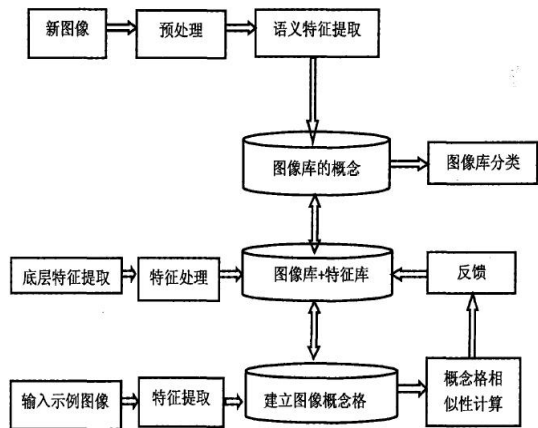


图 2.3 基于形式概念分析的图像检索流程

3 结语

本文重点在于利用形式概念分析给出图像检索的新方法。通过概念格的构造方法，理解形式概念分析，在此基础上建立图像的形式背景，生成概念格来对图像进行检索。由于图像数据库中的图像只有进行相似性计算时才会构成概念格，所以减少了计算时间。另外，如何快速建立图像概念格，更好的计算概念格的相似度是下一步研究的重点。

参考文献：

[1] 赵倩. 基于内容的图像检索若干技术研究 [D]. 上海: 上海大学, 2012.

[2] Tamura H, Yokoya N. Image Database systems: A Survey[J]. Pattern Recognition, 1994. 17(1): 121-128.

[3] 马垣, 曾子维, 迟呈英, 吴建胜等. 形式概念及其新进展 [M], 北京: 科学出版社, 2010, 35—85.

[4] 苏亮. 基于语义图像检索的相关研究[D]. 河北: 华北电力大学, 2011.

[5] JainAK, Duin RP. Statistical pattern recognition: a review Pattern Analysis and Machine Intelligence[J], IEEE Transactions, 2000.