We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter 8

Incorporating Domain Knowledge into Medical Image Mining

Haiwei Pan

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/50207

1. Introduction

Advances in image acquisition and storage technology have led to tremendous growth in very large and detailed image databases [2]. A vast amount of image data is generated in our daily life and each field, such as medical image (CT images, ECT images and MR images etc), satellite images and all kinds of digital photographs. These images involve a great number of useful and implicit information that is difficult for users to discover.

Image mining can automatically discover these implicit information and patterns from the high volume of images and is rapidly gaining attention in the field of data mining. Image mining is more than just an extension of data mining to image domain. It is an interdisciplinary endeavor that draws upon computer vision, image processing, image retrieval, machine learning, artificial intelligence, database and data mining, etc. While some of individual fields in themselves may be quite matured, image mining, to date, is just a growing research focus and is still at an experimental stage.

Broadly speaking, image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the images and between image and other alphanumeric data. For example, in the field of archaeology, many photographs of various archeological sites have been captured and stored as digital images. These images, once mined, may reveal interesting patterns that could shed some lights on the behavior of the people living at that period of time. Clearly, image mining is different from low-level computer vision and image processing techniques. The focus of image mining is in the extraction of patterns from a large collection of images, whereas the focus of computer vision and image processing techniques is in understanding and/or extracting specific features from a single image. While there seems to be some overlap between image mining and content-based retrieval (since both deals with large collection of images), image mining goes beyond the problem of retrieving relevant images. In image mining, the goal is the

discovery of image patterns that are significant in a given collection of images and the related alphanumeric data. Perhaps the most common misconception of image mining is that image mining is yet another term for pattern recognition. While the two fields do share a large number of common functions such as feature extraction, they differ in their fundamental assumptions. In pattern recognition, the objective is to recognize some specific patterns; whereas in image mining, the aim is to generate all significant patterns without prior knowledge of what patterns may exist in the image databases. Another key difference is in the types of patterns examined by the two research fields. In pattern recognition, the patterns are mainly classification patterns. In image mining, the patterns types are more diverse. It could be classification patterns, description patterns, correlation patterns, temporal patterns, and spatial patterns. Finally, pattern recognition deals only with pattern generation and pattern analysis. In image mining, this is only one (albeit an important) aspect of image mining. Image mining deals with all aspects of large image databases which imply that the indexing scheme, the storage of images, and the retrieval of images are all of concerns in an image mining system[3]. A few interesting studies and successful applications involving image mining have been reported. For example, [4] describes the CONQUEST system that combines satellite data with geophysical data to discover patterns in global climate change. The SKICAT system [5] integrates techniques for image processing and data classification in order to identify "sky objects" captured in a very large satellite picture set. A multimedia data mining system prototype MultiMediaMiner [2, 6] uses a data cube structure for mining characteristic, association, and classification rules. However, the system does not use image content to the extent we wanted. In [7], localization of the visual features, their spatial relationships and their motion in time (for video) are presented. A discovering association rules algorithm based on image content from a simple image dataset is presented in [8]. [9]

Research in image mining can be broadly classified into two main directions [3]. The first direction involves domain-specific applications where the focus is to extract the most relevant image features into a form suitable for data mining [10, 11]. The second direction involves general applications where the focus is to generate image patterns that maybe helpful in the understanding of the interaction between high-level human perceptions of images and low level image features [2, 8, 12]. Clustering medical images belongs to the first direction.

Image clustering is unsupervised classification of images into groups. The problem in image clustering is to group a given collection of unlabeled images into meaningful clusters according to the image content without a priori knowledge [13]. The fundamental objective for carrying out image clustering in image mining is to acquire content information the users are interested in from the image group label associated with the image[3]. Image clustering is usually performed in the early stages of the mining process. Feature attributes that have received the most attention for clustering are color, texture and shape. Generally, any of the three, individually or in combination, could be used. There is a wealth of clustering techniques available: hierarchical clustering algorithms, partition-based algorithms, mixture-resolving and mode-seeking algorithms, nearest neighbor clustering,

fuzzy clustering and evolutionary clustering approaches. Once the images have been clustered, a domain expert is needed to examine the images of each cluster to label the abstract concepts denoted by the cluster. Chang et al. use clustering technique in an attempt to detect unauthorized image copying on the World Wide Web [3, 14]. Yu and Zhang present an unsupervised clustering and query approach (also known as ACQ for Automatic Clustering and Query) for large-scale image databases [15]. ACQ does not require the number of clusters to be known a priori and is insensitive to noise. By intelligently applying wavelet transforms on the feature space, this clustering can effectively and efficiently detect clustering of arbitrary shape of high dimensional feature vectors. Kitamoto apply clustering methods such as k-means and the self-organizing map (SOM) for visualizing the distribution of typhoon cloud patterns on a two-dimensional space [3, 11].

Some algorithms used in the medical images [16, 17, 18] are generally for classification. [19, 20, 21] propose some new clustering methods in relational databases. They are not suitable to cluster the medical images because there are important differences between relational databases versus image databases: (1) Absolute versus relative values. In relational databases, the data values are semantically meaningful. For example, one item is milk is well understood. However, in medical image databases, the data values themselves may not be significant unless the domain of medicine supports them. For example, a grey scale value of 46 could appear darker than a grey scale value of 87 if the surrounding context pixels values are all very bright. (2) Spatial information (Independent versus dependent position) [9]. Another important difference between relational databases and medical image databases is that the implicit spatial information is critical for interpretation of image contents but there is no such requirement in relational databases. As a result, image miners try to overcome this problem by extracting position-independent features from images first before attempting to mine useful patterns from the images[3]. (3) Unique versus multiple interpretation. A third important difference deals with image characteristics of having multiple interpretations for the same visual patterns. The traditional data mining algorithm of associating a pattern to a class (interpretation) will not work well here. A new class of discovery algorithms is needed to cater to the special needs in mining useful patterns from images[3].

Association rule mining from medical images is another significant research topic in the field of image mining. Finding these valuable rules is typically done in two steps: discovering frequent itemsets and generating association rules. The second step is rather straightforward, and the first step dominates the processing time, so we explicitly focus this chapter on the first step. A number of efficient association rule mining algorithms have been proposed in the last few years. Among these, the Apriori algorithm by Agrawal, R., Imielinski, T [22][23] has been very influential. Later, many scholars have improved and optimized the Apriori algorithm and have presented new Apriori-like algorithms [24][25][26][27][28]. The Apriori-like algorithms consist of two major procedures: the join procedure and the prune procedure. These algorithms require a huge calculation and a complicated transaction process during the two procedures. Therefore, the mining efficiency of the Apriori-like algorithms is not very good when transaction database is very large.

L.Jaba Sheela proposed the FAR algorithm [29] in 2009. This algorithm transforms a transaction database into a Feature matrix stored in bits. Meanwhile it uses the Boolean vector "relational calculus" method to discover frequent itemsets. This method uses the fast and simple "and calculus" in the Feature matrix to replace the calculations and complicated transactions that deal with large number of itemsets [32]. It scans the database only once and has a good efficiency. But there is a great shortcoming in FAR algorithm. The generating of candidate itemsets relies on the combination of column of feature matrix. When the number of column of feature matrix is very large, that is to say the number of items in transaction database is very large, FAR algorithm will cost much time to do useless calculus.

In this chapter, we firstly quantify the domain knowledge about brain image (especially the brain symmetry), and then incorporate this quantified measurement into the clustering algorithm. Our algorithm contains two parts: (1) clustering regions of interest (ROI) detected from brain image; (2) clustering images based on the similarity of ROI. We apply the method to cluster brain images and present results to demonstrate its usefulness and effectiveness[1]. Secondly, we proposed the GMA (association graph and matrix pruning algorithm) algorithm to solution this problem. Yen and Chen proposed the DLG (Direct Large Itemset Generation Algorithm) algorithm [30][31] based association graph for the first time in 1996. GMA algorithm adopts both association graph and matrix pruning to reduce the generation of candidate itemsets. It also scans database only once and generates frequent itemsets by the "and calculus''. Throughout the chapter we try to provide a general framework to understand these approaches. We believe many of the problems we are facing are likely to appear in other domains. As such this work tries to isolate those problems which we consider will be of most interest to the database community doing research on clustering [9] and association rule mining.

The rest of the chapter is organized as follows: section 2 is pre-processing describing an algorithm to detect objects in medical images [9]. Section 3 presents the medical image clustering algorithm. This section includes four parts: (1) Feature extraction is to extract the most relevant features from the object with the direction of domain knowledge; (2) Object clustering. We define the similarity measurement according to the above features and present OCA algorithm to cluster objects into some groups; (3) Image clustering. We firstly determine weights of objects that appear in images based on term frequency and inverse document frequency, similar to IR. Each image will then be represented by a vector, where a vector contains a set of weights that correspond to the importance of the objects that appear in the image. Finally ICA algorithm is presented to cluster medical images; (4) Experiment results. This part reports the results of our experiments and performance study. Section 4 presents the association rule mining algorithm. This section includes three parts: (1) Basic concept. We define Support, Confidence, Association Graph and Feature Matrix, etc. to describe this problem; (2) GMA algorithm. The GMA algorithm consists of four phases as follows: generating feature matrix and association graph, pruning the feature matrix, selecting and extending by the association graph and generating the set of frequent-k itemsets; (3) Experiment results give the experiment results and analysis. Section 5 introduces the future work. Section 6 concludes the study in this chapter.

2. Preprocessing

Since the images we studied were raw Computerized Tomography (CT) scans that were scanned at different illumination conditions, some of them appeared too bright and some were too dark. We should digitize them to no loss, no compression and 256 gray scale images through special medical scanner. We used CT scan images because this modality is the most used in radiotherapy planning for two main reasons. The first reason is that scanner images contain anatomical information which offers the possibility to plan the direction and the entry points of the radiotherapy rays which have to target the tumor and to avoid some risk organs. The second reason is that CT scan images are obtained using rays, which is the same physical principle as radiotherapy. This is very important because the radiotherapy rays intensity can be computed from the scanner image intensities[33].

In this section, we firstly use progressive water immersion method with guidance of domain knowledge to detect region of interest (ROI) in medical images, then we combine these ROIs with their location, size and other descriptors to form a table for mining.

Water immersion algorithm is considered to be a powerful technique for ROI detection. It works by grouping pixels with similar gradient information. Direct application of water immersion method to the digitized medical images typically produces over-segmentation of the trivial regions. Instead, we propose a progressive water immersion algorithm with guidance of domain knowledge to cope with this situation [34]. Details of the algorithm follow.

First, a N×N window is used to locate the local optimal points in the image. For each segmented patches, we place the center of the window over each pixel in the patches. If the grey level of the central pixel is optimal with respect to all the other pixels in the window, we say that the central pixel is a local optimum; otherwise, the window will move to be centered at another pixel to continue the search for all local optimal points. At the end of this phase, all the optimum is marked and they will be treated as the starting seeds for water immersion method. One advantage of using the sliding window approach is that with the appropriate window size, it is possible to eliminate a large amount of optimal points that correspond to the light and dark reflection regions thus removing false detection. This is because the grey level of the optimal points corresponding to the light and dark reflection patches are generally lower and higher than that of potential ROI. Given that the distances between the optimal points of the light and dark reflection patches and the nearest optimal points of the neighboring ROI are generally less than that between two touching ROI, it is possible to set the window size in such a way that these false optimal points are 'absorbed' by the neighboring ROI optimal points while the true optimal points are not affected.

Having identified the true optimal points, water immersion process starts from these detected points and progressively floods its neighboring pixels. The neighboring pixels are defined to be the 8-direction neighbors. These neighbors are placed in a growing queue structure sorted in descending order of the grey level of the pixels. The lowest and highest grey pixel in the growing queue will be 'immersed' first but respectively and it is marked as

belonging to the same region label as the current seed. The marked pixel is then removed from the growing queue. All neighboring pixels whose grey level is lower or higher than the marked pixel are added to the growing queue. This immersion process continues until the growing queue is empty [9].

Unfortunately, simple application of the water immersion technique has the tendency of over-flooding. To overcome this problem [9], we firstly give some definitions.

1. We partition all pixels in pixel set P into m blocks. Pixels in the same block have the same grey level and pixels in the different blocks have the different grey level. Let $G(P)=\{g_1, g_2, \ldots, g_m\}$ be P's grey-scale (GS) set if G(P) is an ascending sort set of g_1', g_2', …, g_m' and g_i' is grey level of pixels in the i^{th} block, where g_i (i=1,…,m) is the i^{th} GS, g_1' and g_m' are P's minimum and maximum GS respectively. The GS of pixel p_i is denoted as $g(p_i)$ [33].
2. We call $g_{mean}(P)$ the mean GS if

$$g_{mean}(P)= \sum_{i=1}^{|P|} g(p_i)/|P|.$$

3. For any P and distance function $DisA=|g_k - g_{mean}(P)|$, mid-value GS is a middle value in the GS set that minimizes DisA. Mid-value GS set is a set of mid-value GS [33].
4. For any P, if

a. Mid-value GS set includes one element g_{mid}, and g_s is the minimum value between g_{mean} and g_{mid};
b. Mid GS set includes two elements, g_s is the minimum value between these two values; g_s is called Benchmark GS and another one is denoted as g_s' [33].

5. For pixel set P, let

$g^{(l)}=\{g_i \mid g_1 \le g_i \le g_1+|g_1-g_s|/2\}$ be low bound GS;
$g^{(h)}=\{g_i \mid g_m-|g_m-g_s'|/2 \le g_i \le g_m\}$ be high bound GS;
$g^{(b)}= g^{(l)} \cup g^{(h)}$ be bound GS; [33]

Our progressive water immersion ignores all those pixels whose grey level doesn't belong to the bound GS. Bound GS is defined with guidance of domain knowledge that describes the degree of dark and light. So the optimality of point in a certain region is defined as follows [9]:

$$\text{optimality}=\begin{cases} \text{maximal grey level, if point belongs to high bound pixel set ;} \\ \text{minimal grey level, if point belongs to low bound pixel set ;} \end{cases}$$

The pseudo-codes for the progressive water immersion algorithm are given as follows.

Input: medical image
Output: objects in this medical image

```
 WHILE(image scan not finished)
{
    IF(the first scan)
       initialize slide window;
    ELSE
       relocate initial position of the slide window;
    WHILE (1)
    {
       compare pixel grey value in the window and find the optimal point;
            IF (the optimal point is the center of the window)
                this point is stored as seed;
                Break;
           ELSE
            move the slide window and make the optimal point as the center of window;
    }
}
Count=1;
Con_th= seed-to-pixel contrast threshold;
FOR(each seed)
{
       FOR (each 8-directional neighbor pixel of the seed)
          IF (absolute value of seed-to-neighbor pixel contrast is less than Con_th)
          {
             push this pixel into queue and mark this pixel;
             Count++;
          }
       WHILE (Count!=0)
       {
          sort pixel in queue in descent order according to the grey value;
           pop the last pixel from queue;
           Count--;
          FOR (each 8-directional neighbor pixel of the last pixel)
            IF(absolute value of seed-to- neighbor pixel contrast is less than Con_th)
            {
              push this pixel into queue and mark this pixel;
              Count++;
            }
       }
}
```

After the above process, all the ROIs are detected and we will call them objects later, see figure 1. Next, images with many different objects are represented by transactions and we use a table to describe these transactions, see table 1 [9].

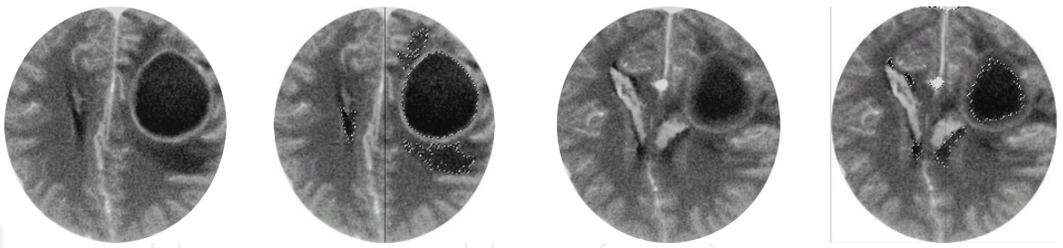


Figure 1. Figure (a) and (c) are two original abnormal brain images. Progressive water immersion algorithm is used to mark the objects with dotted line in figure (b) and (d).

In table 1, the first column is the medical image id. The second column is the objects in each image. The other columns are the features extracted from the object.

Image ID	Object ID	feature$_1$	…	feature $_n$
IM_1	O_1	feature$_1$_v	…	feature$_n$_v
IM_1	O_2	…	…	…
IM_1	O_3	…		
IM_2	O_1	…		
IM_2	O_2	…		
…	…	…		
IM_n	O_n	…	…	…

Table 1. Images are modeled by transactions

3. Medical image clustering

By applying progressive water immersion algorithm, we segment images into objects. Let IM={IM_1, IM_2, …, IM_N} be a image set. After the above algorithm, Each image IM_j contains k objects R_{j1}, R_{j2}, …, R_{jk}. For different image, k may be not equal. Let the total of objects in IM be M, then we denote the object set as R={ R_1, R_2, …, R_M}.

3.1. Feature extraction

For each extracted object R_i, we need to extract relevant information for features mining to take place. The domain knowledge of the brain image characteristics indicates that the normal persons have nearly the same brain structure that is evident to be bilateral symmetry. That is, the distribution of density in the left hemisphere of the brain is almost identical with the right, see figure 2 [9]. The pathological regions result in irregularly shaped grey level distribution in the CT scan images and destroy the symmetry, see figure 1. At this point, the following relevant features are able to provide sufficient discriminative power to cluster objects into different groups: (1) grey level of the object of interest; (2) area of the object of interest; (3) location of the object of interest; (4) elongation of the object of interest; (5) direction of the object of interest; and (6) symmetry of the object of interest.

Figure 2. Normal person's brain image

With the above method, the object extracted from brain image is either brightness or darkness. So we define grey level as GL= 0 (brightness), or 1 (darkness). The second feature we have found useful relate to the size of the region in the form of the area of the region. Area of the region of interest is defined as the total number of pixels within the region, up to and including the boundary pixels. The location of an object of interest is defined as a ratio. The coordinates of the centroid of the object [9] are computed with the formula (1).

$$\bar{x} = \frac{1}{k}\sum_{j=1}^{k} x_j \qquad \bar{y} = \frac{1}{k}\sum_{j=1}^{k} y_j \tag{1}$$

k is the number of the pixels of the object and the location is ($\bar{x}$ /|x|, $\bar{y}$ /|y|).

The fourth feature is the elongation of an object which is defined as the ratio of the width of the minor axis to the length of the major axis. This ratio is computed as the minor axis width distance divided by the major axis length distance, giving a value between 0 and 1. If the ratio is equal to 1, the object is roughly a square or is circular in shaped. As the ratio decreases from 1, the object becomes more elongated. Major axis is the longest line that can be drawn through the object. The two end points of the major axis are found by selecting the pairs of boundary pixels with the maximum distance between them. This maximum distance is also known as the major axis length. Similarly, the minor axis is defined as the line that it is perpendicular with respect to the major axis and the length between the two end points of the line intersected with the object is the longest which is called the width of the minor axis. The ratio, elongation, is a measure of the degree of elongation of an object.

$$elongation = len_{major} / len_{minor} \tag{2}$$

where len_{major} is the major axis length of the object and len_{minor} is the minor axis length of the object. We establish a common coordinate with brain midline as y and perpendicular line going through the midpoint of midline as x. With the major axis of the object, we define the next feature, direction of the object, as the inclination θ of the major axis and x positive direction and $\theta \in [0,180]$ [9].

Before introducing the final feature, we will give some definitions.

6. Pixel set of IM_P is defined as $P=\{p_i | p_i$ is the pixel with coordinate (x_i, y_i) in the image $IM_P\}$, P(L) and P(R) are pixel set of IM_P(L) and IM_P(R) respectively [33].
7. For any $p_{li} \in P(L)$, $p_{rj} \in P(R)$, they are symmetric pixel if the line between p_{li} and p_{rj} is halved vertically by brain midline. They are denoted as p_{li} and p_{ri} below [33].

8. $\Delta g(P)$ is IM_P's difference set if for any symmetrical pixel p_{li} and p_{ri}, $\Delta g(P)=\{ \Delta g_i | \Delta g_i=g(p_{li})-g(p_{ri}), i=1,2,\ldots,|p|/2 \}$, where $g(p_{li})$ and $g(p_{ri})$ are grey level of these two pixels respectively.

Now we define the sixth feature as following.

9. An object of interest R_{pk} is symmetrical if for IM_P and R_{p1}, R_{p2}, …, R_{pk}, $P' = P - \sum_{i=1}^{k} P(R_{pi})$, the mean grey level of $\Delta g(R_{pk})$, mean($\Delta g(R_{pk})$), and the mean grey level of $\Delta g(P')$, mean($\Delta g(P')$), satisfy the following condition:

$$|\text{mean}(\Delta g(R_{pk})) - \text{mean}(\Delta g(P'))| < \varepsilon.$$

Theorem 1: If for IM_P, an object of interest R_{pk} is symmetrical, then there must exist another object of interest R_{pt} that satisfies the following conditions: (1) R_{pk} and R_{pt} must lie in different side of the midline; (2) R_{pk} and R_{pt} are either two different objects or the same one.

Proof: According to the definition of symmetric pixel and object symmetry, condition 1 is evident. For the second condition, it is also evident for R_{pk} and R_{pt} to be two different objects. If R_{pk} bestrides two hemispheres and is symmetrical, it satisfies the definition of object symmetry. Actually, R_{pt} is the part of R_{pk} that lies in the other side of the brain midline. Both of them belong to the same object.

3.2. Object clustering

Now, we would like to determine similarity between two objects based on these features. For an object R_i, a vector $V_i(v_{i,1}, v_{i,2}, \ldots, v_{i,k})$ is constructed and similarity between object R_i and object R_j is as follows [9]:

$$\text{Sim}^R(R_i,R_j) = \Delta v_{ij,1} * \left(\sum_{h=2}^{k-1} v_{i,h} * v_{j,h}\right) / \left(\sqrt{\sum_{h=2}^{k-1} v_{i,h}^2} * \sqrt{\sum_{h=2}^{k-1} v_{j,h}^2}\right) \quad (3)$$

According to domain knowledge, we let $v_{ij,1}$ be GL. Two objects are not possible to be similar if one is very darkness and the other is very brightness. So we let $\Delta v_{i,1}$ be 1 if $v_{i,1} = v_{j,1}$, or 0 if $v_{i,1} \neq v_{j,1}$ in formula (3) [9]. Another similar rule is that if two objects R_i and R_j satisfy the second condition in theorem 1, they will be grouped into the same cluster. This rule is prior to the above similarity function.

It is difficult for us to know the number of the clusters in advance. So we use DBscan algorithm to group these objects. We use a non-negative threshold T_R to construct object clusters. If similarity between two objects is smaller than T_R, then the two objects can be in the same group [9].

10. For an object R_j, its ε-neighborhood, denoted by ε-$N(R_j)$, is defined by:

$$\varepsilon\text{-}N(R_j) = \{X \in R \mid \text{Sim}^R(R_j,X) \leq T_R \}$$

11. R_i is core object (c-object) if its ε-N(R_i) involves no less than MP objects, that is, |ε-N(R_i)|≥MP.
12. R_i is directly density reachable from R_j if $R_i \in$ ε-N(R_j) and |ε-N(R_i)|≥MP (R_i is c-object).
13. R_1 is density reachable from R_k if there is a chain of objects R_1, R_2, …, R_k, any object R_{i+1} is directly density reachable from R_i.
14. R_i is density connected to R_j if there is an object R_k that is density reachable from not only R_i but also R_j.

Density-connectivity is a symmetric relation. For density reachable points, the relation of density-connectivity is also reflexive. The key idea is that for each object of a cluster the neighborhood of a given radius has to contain at least a minimum number of objects, i.e. the density in the neighborhood has to exceed some threshold. The shape of a neighborhood is determined by the similarity function. To find a cluster, OCA starts with an arbitrary object R and retrieves all objects density-reachable from R. If R is a core object, this procedure yields a cluster. If R is a border object, no objects are density-reachable from R and OCA visits the next object of the dataset.

If two clusters C_1 and C_2 are very close to each other, it might happen that some object R belongs to both C_1 and C_2. Then R must be a border object in both clusters because otherwise C_1 would be equal to C_2 since we use global parameters. In this case, object R will be assigned to the cluster discovered first. Except from these rare situations, the result of our algorithm is independent of the order in which the objects of the database are visited.

In the following, we present the Object Clustering Algorithm (OCA) [9].

Object Clustering Algorithm (OCA):
Input: object set R, T_R and MP
Output: p clusters

```
1. Assume that the size of object set R is M and examine ε-N(R);
2. For k = 1 to M {
3.   If (Ri is unclassified) Then
4.     If (ε-N(Ri) involves more than MP objects) Then {
5.       mark Ri as the core object;
6.       Cluster_Expanding(Ri);
7.     else mark Ri classified;
8.     }
9. }
```

The function *Cluster_Expanding()* is most important for discovering arbitrary shape clusters. It is presented bellow [9].

Cluster_Expanding(Ri)

1. While(all core objects) {
2. mark Ri *classified*;
3. cluster all density reachable objects;
4. record all the core objects;
5. mark those non-core objects as *classified*;
6. }

The average run time complexity of this algorithm is O(MlogM) with spatial index. Otherwise, it is $O(M^2)$.

3.3. Image clustering

The OCA algorithm clusters these M objects into p groups which we denoted as RC={C_1, C_2, …, C_p}. Image clustering is based on image similarity. To calculate image similarity, we construct a vector $W_i(w_{1,i}, w_{2,i}, …, w_{p,i})$ for an image IM_i. $w_{p,i}$ is the weight of object cluster C_p in image i. In this vector we keep the weight of each group. Thus, the size of vector W_i is same as the total number of object clusters (=p). It is possible that the weight of a group may be zero. This is because no object of an image may be a participant of that group during clustering. To determine an image vector, we adopt the idea from the area of information retrieval. Here, images correspond to documents, and object clusters correspond to terms (keywords) [9].

15. Let $CIM_{j,i}$ be the times of objects of cluster C_j in image IM_i, IMC_j be the number of images in which objects of cluster C_j appear. We define

$$iIMC_j = \log (N/IMC_j) \tag{4}$$

where N is the size of the image set [9].

IMC_j is used to evaluate the function of C_j to measure the similarity of two images. For $iIMC_j$, the higher its value is, the more impact C_i has to distinguish two different images.

16. For a vector $W_i(w_{1,i}, w_{2,i}, …, w_{p,i})$, we define

$$w_{j,i} = CIM_{j,i} * iIMC_j \tag{5}$$

After computing image vectors, we get similarity between any two images using cosine similarity:

$$Sim^{IM}(IM_i,IM_j) = \left(\sum_{h=1}^{p} w_{h,i} * w_{h,j}\right) / \left(\sqrt{\sum_{h=1}^{p} w_{h,i}^2} * \sqrt{\sum_{h=1}^{p} w_{h,j}^2}\right) \tag{6}$$

The higher the value of Sim^{IM} is, the more similar the two images are [9].

When two microclusters, each of which includes more than two images, are to be grouped into a new cluster, we redefine the similar function to measure the similarity of two microclusters.

$$\mathrm{Sim}^{MC}(MC_i, MC_j) = \left(\sum_{h=1}^{p} \overline{w}_{h,i} * \overline{w}_{h,j}\right) / \left(\sqrt{\sum_{h=1}^{p} \overline{w}_{h,i}^{2}} * \sqrt{\sum_{h=1}^{p} \overline{w}_{h,j}^{2}}\right) \quad (7)$$

where $\overline{W}_i(\overline{w}_{1,i}, \overline{w}_{2,i}, ..., \overline{w}_{p,i})$ is the centroid of all vectors within microcluster MC_i.

The algorithm will stop when the images are clustered into k group.

Image Clustering Algorithm (ICA):
Input: Image set IM and the number of clusters k;
Output: k clusters

1. Each element of IM is regarded as an atomic cluster and compute Sim^{IM};
2. Find the biggest Sim^{IM} and Amalgamate to form a new cluster;
3. While (the number of clusters is not equal to k) {
4. If all R_ts in one image are symmetrical
5. Then cluster this image to *special* cluster;
6. Compute Sim^{MC};
7. Cluster the sub-clusters or images; }

The average run time complexity of this algorithm for the worst case is $O(n^2)$. The clustering algorithm starts with each input image as a separate cluster, and at each successive step merges the closest pair of clusters. In order to compute the distance between a pair of clusters, for each cluster, c representative images are stored. These are determined by first choosing c well scattered image within the cluster, and then shrinking them toward the mean of the cluster by a fraction α. The distance between two clusters is then the distance between the closest pair of representative images - one belonging to each of the two clusters. Thus, only the representative images of a cluster are used to compute its distance from other clusters.

The c representative images attempt to capture the physical shape and geometry of the cluster. Furthermore, shrinking the scattered images toward the mean by a factor α gets rid of surface abnormalities and mitigates the effects of outliers. The reason for this is that outliers typically will be further away from the cluster center, and as a result, the shrinking would cause outliers to move more toward the center while the remaining representative images would experience minimal shifts. The larger movements in the outliers would thus reduce their ability to cause the wrong clusters to be merged. The parameter α can also be used to control the shapes of clusters. A smaller value of α shrinks the scattered images very little and thus favors elongated clusters. On the other hand, with larger values of α, the scattered images get located closer to the mean, and clusters tend to be more compact [9].

3.4. Experiment results

The main reason why we study on real brain CT images instead of any simulative data is to avoid insignificance and uninterestingness and the reliability of the discovered knowledge [9]. On the other hand, it is because brain tissue is human's advanced nerve center, its function is particularly important. The disease affecting the brain has received much attention in the domain of medicine. In China, about 40,000 to 60,000 persons suffer from

brain tumor every year. Especially during these years, the incidence of brain disease (especially brain tumor) has increased significantly. Therefore, the early diagnosis of brain diseases is becoming more and more crucial and is directly working on patients' treatment [33]. To have access to real medical images is a very difficult undertaking due to legally privacy issues and management of hospital. But with some specialists' help and support, we got 618 precious images and their corresponding diagnosis records which, for simplicity, we generalized to normal(N) and abnormal(A) [9].

Our algorithm is written in Visual C++ and compiled by Microsoft Visual Studio 6.0. All of experiments are performed on Acer computer using a 2.8GHZ Intel PC, 512MB of RAM, and 1024MB of virtual memory. The operating system in use was the Microsoft Windows XP.

To measure the quality of a cluster, we use precision, recall, and E measure. Recall is the ratio of relevant images to total images for a given category. Precision is the ratio of relevant images to images that appear in a cluster for a given category.

$$\text{Precision(P)}= \frac{\text{number of images correctly classified into each class}}{\text{number of total images}}$$

$$\text{Recall(R)}= \frac{\text{number of images correctly classified into one certain class}}{\text{number of images in one certain class}}$$

E measure is defined as follows:

$$E(p,r) = 1 - \frac{2}{1/p + 1/r}$$

where p and r are the Precision and Recall of a cluster. Note that E (p, r) is simply one minus harmonic mean of the precision and recall; E (p, r) ranges from 0 to 1 where E (p, r) =0 corresponds to perfect precision and recall, and E (p, r) corresponds to zero precision and recall. Thus, the smaller the E measure values the better the quality of a cluster [9].

The existing image clustering methods generally include two parts: (1) extract related features from image to form feature vector; (2) use distance function as the image similarity measurement to cluster images. Therefore, we design an image feature-based method as the comparison with our ROI-based method (ICA). Firstly, two related features we extract from medical images are asymmetry and grey level mean difference. Then, k-means algorithm is used to cluster these medical images. In the experiment, we sample five patients' images each time to test two clustering algorithms. Each patient has an image sequence. All the images in the same sequence, which are similar since they belong to the same patient, should be clustered into one group. In this way, we sample five times from the medical image dataset and get the average value of precision, recall and E value for each time, see figure 3-5. It is obviously that ICA is better than the image feature-based method because the precision and recall of ICA is higher than the image feature-based method and E value is lower.

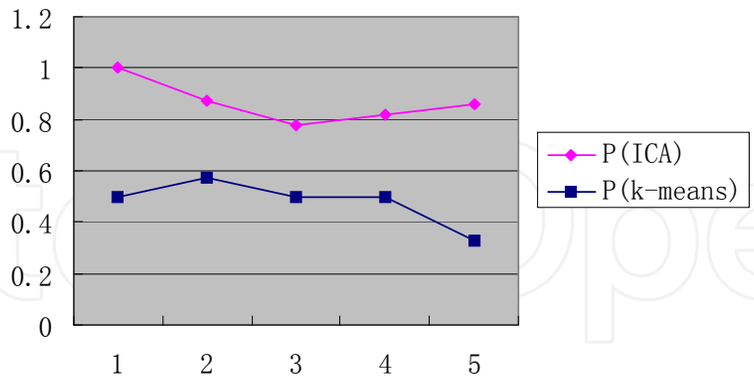


Figure 3. Precision for different number of sampling

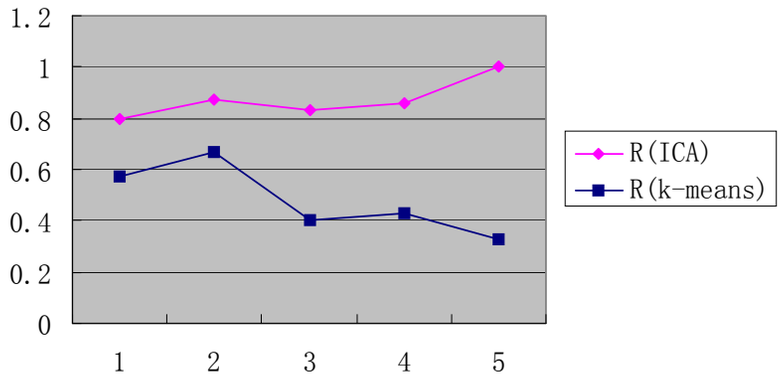


Figure 4. Recall for different number of sampling

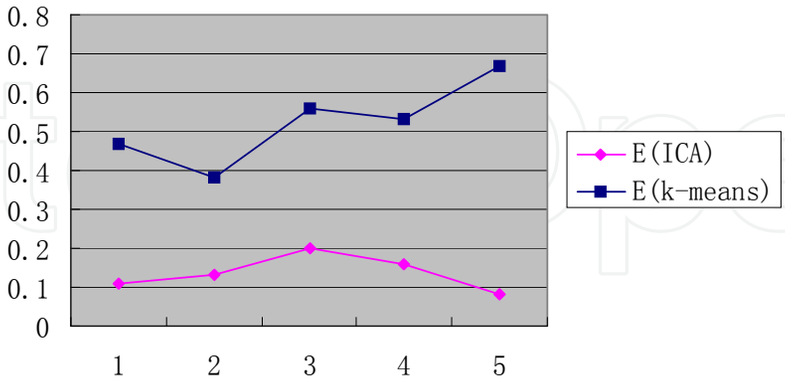
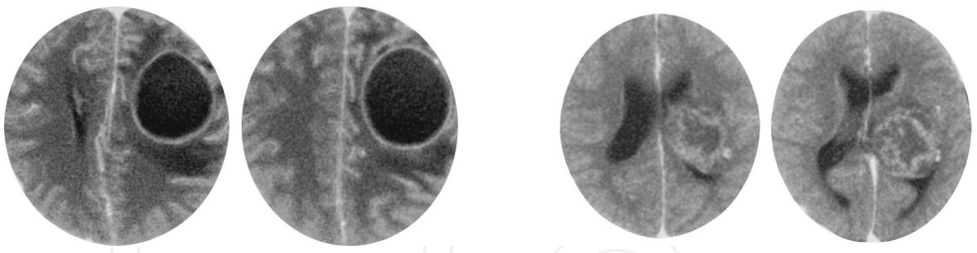


Figure 5. E measure for different number of sampling

In figure 3-5, x axis represents different sample time and the y axis represents p, r and E respectively. We have observed that precision and recall are higher for medical images.

Figure 6 shows an instance of our clustering algorithm.



(a) (b)

Figure 6. (a) and (b) show two images of two different clusters

In any case fairly large medical data sets exist but they are not available to us. Also, it would be interesting to apply these ideas in other domains where large complex data sets are available [9].

4. Association rule mining

In this section, we proposed the improved algorithm that generating candidate itemsets based association graph and matrix algorithm (GMA). That is, the GMA algorithm's candidate itemsets is the intersection of DLG and FAR algorithm's candidate itemsets. Experiments show that, GMA algorithm reduced the candidate itemsets generation greatly and had higher efficiency compared with other algorithms. The DLG algorithm's main idea is to construct a direct edge for every itemset of frequent-2 itemsets L_2. So L_2 can be mapped to a digraph, association graph. DLG algorithm uses the information of association graph to mine the frequent-k itemsets, where k is an integer. The FAR algorithm's main idea is to map the transaction database to a matrix with elements values of '0' and '1'. In order to mine frequent itemsets fast, it deletes some column and row to prune the matrix. Moreover, it must consider column of the matrix first, then consider the row of the matrix. The FAR algorithm also needs to generate the candidate itemsets. The candidate-k itemsets generate from k-vectors combination of columns of matrix. If the feature matrix has n columns, the number of candidate-k itemsets is C_n^k. Then it do "and calculus" for each combination of k-vectors. If the sum of element values in the "and" calculation result is not smaller than the minimum support, the k-itemsets corresponding to this combination of k-vectors are the frequent k-itemsets and are added to the set of frequent k-itemsets Lk [32]. DLG algorithm relies on the information of association graph to generate candidate itemsets, but when the transaction database is very large, the association graph is large, and the number of candidate itemsets also is very large. FAR algorithm generates the candidate-k itemsets by arbitrary k-vectors combination, that is C_n^k. The column of matrix is very large normally, so the number of candidate itemsets is also large.

4.1. Basic concept

A formal statement of the association rule is shown in Definition 1, 2 and 3.

17. Let I = {I_1, I_2, …, I_m} be a set of m distinct attributes, also called literals. Let D be a database, where each record (tuple) T has a unique identifier, and contains a set of items

such that $T \subseteq I$. An association rule is an implication of the form $X \Rightarrow Y$, where $X, Y \subset I$, are sets of items called itemsets, and $X \cap Y = \phi$. Here, X is called antecedent, and Y consequent [3].

Two important measures for association rules, support (s) and confidence (α), can be defined as follows.

18. The support (s) of an association rule is the ratio (in percent) of the records that contain $X \cup Y$ to the total number of records in the database.
19. For a given number of records, confidence (α) is the ratio (in percent) of the number of records that contain $X \cup Y$ to the number of records that contain X.

Mining of association rules from a database consists of finding all rules that meet the user-specified threshold support *s* and confidence *a*.

20. Let D be a transaction database. L_k is the set of frequent-k itemsets, where k is an integer, for each itemset of L_k, it occurs with a frequency that is greater than or equal to the user-specified threshold support, s.
21. Association Graph.

For a directed graph G = {V, E}. V(G) is the collection of vertex in G. E(G) is the collection of edge in G. $<V_i, V_j>$ is a direct edge that from vertex of V_i to $V_j\{i < j\}$. Suppose that V_i, V_j mapped to item O_i,O_j in transaction database DB. Moreover, G can meet the following conditions:

$$V(G) = \{V_i \mid O_i \in L_1\},$$

$$E(G) = \{<V_i, V_j> \mid \{O_iO_j\} \in L_2, i<j\}$$

G is the association graph of the transactions database .

Property 1. Suppose the directed graph G is the association graph of image database D, the number of edge in G is equal to the number of frequent-2 itemsets $|L_2|$ in image database D.

22. Association Graph Extend and Join

Suppose $X = \{O_1O_2 \ldots O_k\}$ is a k-itemset of the database, if the association graph G of database has a direct edge $e = <V_k,V_j>$, X can be extended and joined to $X' = \{O_1O_2 \ldots O_k,O_j\}$.

Property 2. Suppose $X = \{O_1O_2 \ldots O_k\}$ is a k-itemset of the database, $X' = \{O_1O_2 \ldots O_k,O_j\}$ is a (k+1)-itemset which extended from k-itemset X by the association graph G. The number of X′ is equal to the outdegree of the vertex V_k in G.

23. Feature Matrix

For a transaction database D, T is the transactions and O is the set of item in D. R is the binary relation from $T = \{T_1T_2,\ldots,T_m\}$ to $O = \{O_1,O_2,\ldots,O_n\}$, R: if the item O_i occur in transaction T_i, set $R(T_i,O_j) = 1$; else set $R(T_i,O_j) = 0$. FM is the feature matrix of the transaction database D. We denote the feature matrix as $FM = (r_{ij})_{m \times n}$, $(r_{ij}) = R(T_i, O_j)$.

For the sake of convenience, we denote the feature matrix $FM_{m\times n}$ and item O_j as $A_{m\times n}$ and I_j respectively.

4.2. GMA algorithm

This section proposed the GMA algorithm to mine the association rules for ROI in medical images database. In this section, we will give the algorithm details. In general, the GMA algorithm consists of four phases as follows: generating feature matrix and association graph, pruning the feature matrix, selecting and extending by the association graph, generating the set of frequent-k itemsets Lk(k>2).

4.2.1. Generating Feature Matrix and Association Graph

In the first phase, GMA algorithm has two steps. One is to transform the medical image transaction database into the feature matrix and get the set of frequent-1 itemsets L_1. The other is to get the set of frequent-2 itemsets L_2 and the association graph of medical image database.

Suppose that the mined medical image transaction database is D, with D having m images and n categories of ROI. First, GMA algorithm transforms the medical image transaction database into the feature matrix. Let $T=\{T_1,T_2,\ldots,T_m\}$ be the set of transactions and $I=\{I_1,I_2,\ldots,I_n\}$be the set of items. So the D can be mapped to a matrix $A_{m\times n}$ relative to ROI which has m rows and n columns. Scanning the image transaction database D, if item I_j is in transaction T_i , where $1\le j\le n, 1\le i\le m$, the element value of A_{ij} is '1', otherwise the value of A_{ij} is ' 0' [32].

After transforming, GMA algorithm scans the ROI matrix, computes the supports of all items, and generates the set of frequent-1 itemsets L_1. The support number I_j.support of item I_j is the number of '1' in the jth column of the feature matrix $A_{m\times n}$. If I_j.support is not smaller than the minimum support, I_j.support≥min_sup, I_j will be added to the set of frequent-1 itemsets L_1. Otherwise the column of the jth will be deleted from the feature matrix.

The phase of Generating feature matrix can be detailed in Algorithm 1.

Algorithm 1: **Feature Matrix Construction**
Input: DB, min_sup.
Output: Feature Matrix A, L_1.
Method：

```
for i = 1 to m
  for j = 1 to n
    set Aij = 0;
Scan the DB;
for j = 1 to n
   for i = 1 to m
    if item j in the ith transaction do begin
      set Aij to 1;
    end
```

```
end
L1= Φ;
Scan Feature Matrix A;
for i = 1 to n do
    denote the 1's counts of the column of Aj as sum(Aj);
    if sum(Aj) ≥ min_sup
       L1=L1∪{Ij};
     else delete Aj from A;
end
```

The set of frequent-2 itemsets generates to construct the association graph. For each itemset I_i of L_1, a node is allocated for item I_i and I_i.link = NULL. For every combination of I_i and I_j ($i<j$) in L_1, the ith and jth column of feature matrix are two vectors, the "and" relational calculus $I_i \wedge I_j$ is done. If the sum of element values in the "and" calculation result is not smaller than the minimum support number min_sup, $sum(I_i \wedge I_j) \geq min\_sup$, a directed edge from item I_i to item I_j is constructed and add itemset $\{I_iI_j\}$ to the set of frequent-2 itemset L_2. So after generating the L_2, we can get the association graph, see Algorithm 2 [32].

Algorithm 2: Association Graph Construction
Input: L_1, min_sup.
Output: L_2, Association Graph.
Method:

```
if L1 ≠Φ, do
   allocate a node for every item Ii of L1, and do
   Item[Ii].link = NULL;  then do
      produce every 2-vectors Ai, Aj(i＜j) combination
        B=Ai∧Aj;
      if the 1 counts of vector B, sum(B) ≥ min_sup do
                allocate a node p;
                p.link = Item[Ii].link; p.Item = Item[Ij];
                Item[Ii].link = p; L2 = L2 ∪{Ii,Ij};
            end
end
```

4.2.2. Pruning matrix

In order to introduce the pruning principle clearly, we give two properties and their proof in the following.

Proposition 1. Let X is a k-itemset, $|L_{k-1}(j)|$ presents the number of items 'j' in all frequent (k-1)-itemsets of the frequent set L_{k-1}. There is an item j in X. If $|L_{k-1}(j)| < k-1$, itemset X is not a frequent itemset.

Proposition 2. For each row vector A_i in the feature matrix of the transaction database D, If the sum of '1' in a row vector A_i is smaller than k, it is not necessary for A_i attending calculus of the k- supports . [32]

```
Algorithm3:OptMatrix( )
Input: L_{k-1}, Feature Matrix A_{m×n}.
Output: Matrix B_{p×q}.
Method:
Scan the matrix A;
for j = 1 to n
        if |L_{k-1}(A_j)| < k-1
          Delete the column A_j of Matrix A;
    for i = 1 to m
    Compute the 1'counts of A_m;
        if sum(A_m) < k
          Delete the row A_m of Matrix;
   for i = 1 to p
     for j = 1 to q
          Output the corresponding matrix B_ij;
```

Pruning the matrix means deleting some rows and columns from it. According to the proposition 1 and proposition 2, we can get the pruning principle. The pruning principle can be described as Algorithm 3. First, the column of the feature matrix is pruned. This is described in detail as: Let I′ be the set of all items in the frequent set L_{k-1}, where k>2. Compute all $|L_{k-1}(j)|$ where $j \in I'$, and delete the column of correspondence item j if $|L_{k-1}(j)|$ is smaller than k-1. Second recompute the sum of the element values in each row in the feature matrix. These rows of the feature matrix whose sum of element values is smaller than k are deleted from this matrix. [32]

4.2.3. Selecting and extending by association graph

GMA algorithm generates the candidate-k itemsets depending on selecting and extending the frequent-(k-1) itemset by the association graph. For a (k-1)-itemset of L_{k-1}, if it do not contain the item j which the corresponding column of feature matrix was deleted by pruning in the (k-1)th pass. GMA extends it to a k-itemset as a candidate-k itemset.

In order to generate candidate-k itemsets, GMA need to consider all itemsets of L_{k-1}. However, this procedure performed following the feature matrix pruning procedure. That is to say, when GMA mine the frequent-k itemsets, it must do the two steps, one is the feature matrix pruning, the other is selecting and extending frequent-(k-1) itemsets to generate the candidate-k itemsets. If there is a column of matrix has been deleted by optimizing matrix, we will not consider the itemset of L_{k-1} which contains the corresponding of item. Otherwise, for each itemset $\{I_1, I_2, \ldots, I_{k-1}\}$ of L_{k-1}, finding edges that from vertex I_{k-1} to other vertex in association graph. If there is an edge from vertex I_{k-1} to vertex u, the itemset $\{I_1, I_2, \ldots, I_{k-1}, u\}$ is a candidate-k itemset. Not that all (k-1)-itemset needs to extend, this idea can be described by Algorithm 4.

Algorithm 4: SEJ( ) Select to Extend and Join
Input: L_{k-1}, Association Graph.
Output: Candidate-k itemset C_k.

```
Ck =Φ;
Scan the Lk-1;
for i = 1 to | Lk-1|
    if {I1, I2, …, Ik-1}∈Lk-1&& {I1, I2, …, Ik-1} not contain item Id
    which the corresponding column of matrix was deleted
    do begin
      pointer = Item[Ik-1].link;
      while (pointer ≠ NULL) do begin
          u = pointer.Item;
          Ck = Ck ∪ {I1, I2, …, Ik-1,Iu};
          pointer = pointer.link;
    end
end
```

4.2.4. Generating Frequent-k Itemsets Lk(k>2)

The most important phase in GMA algorithm is to generate the set of frequent-k itemsets L_k (k>2). In order to find the set of frequent-k (k>2) itemset, GMA algorithm firstly optimizes the matrix. Afterwards it generates the candidate-k itemsets using the information of association graph and matrix pruning. At last it verifies whether the candidate-k itemset is a frequent-k itemset. The details of above procedures are described as follows.

Proposition 3. $|L_k|$ presents the number of k-itemsets in the frequent set L_k. If $|L_k|$ is smaller than k+1, the maximum length frequent itemsets is k.

Most of algorithm for mining frequent-k itemsets's terminal condition is that L_{k-1} is null. However, GMA algorithm terminal condition is described by Property 3. that is to say, when GMA get the L_k, it examine the number of L_k, if $|L_k|$ is smaller than k+1, GMA algorithm terminate to perform the the k+1th mining.

In order to generate the set of frequent-k itemsets L_k, GMA must obtain the candidate-k itemsets, then to verify by the "and calculus" operation. We can get the candidate-k itemsets by the Algorithm 4. GMA algorithm is an iterative to mine the frequent itemsets.

GMA algorithm does the "and" relational calculus to verify whether the candidate-k itemset is a frequent-k itemset of L_k. The "and" relational calculus is for combination of k vectors which corresponding to the candidate-k itemset $\{I_1, I_2, \ldots, I_{k-1}, u\}$. Then make k=k+1, do this again to find all frequent-k itemsets, see Algorithm 5.

Algorithm 5: Frequent-k (k>2)Itemset Generation
Input: L_{k-1}, min_sup.
Output: L_k.
Method:

```
while (|Lk-1| ≥ k) do begin
Lk=Φ;
    OptMatrix(A );
    SEJ(Lk-1, Association Graph);
for i = 1 to |Ck|
if Ck = {I1, I2, …, Ik-1,u}&&
  sum(A1∧A2∧…∧Ak-1∧Au) ≥ min_sup do
    Lk = Lk∪{I1, I2, …, Ik-1,Iu};
 end
k = k+1;
end
```

4.3. Experiment results

In order to appraise the performance of the GMA algorithm, we conducted an experiment using the DLG algorithm, the FAR algorithm and GMA algorithm. These algorithms were implemented in C and tested on Windows XP SP3 Operating system, Mobile Intel Pentium 4 1.89GHz CPU，1024MB DDR RAM，VC++6.0 compiler platform. The test database T10I4D100K was generated synthetically by an algorithm designed by the IBM Quest project. The number of items N is set to 1000; |D| is the number of transactions; |T| is the averages size of transactions, and |I| is the average size of the maximum frequent itemsets. Fig. 7 presents the experimental results for different values of minimum supports.

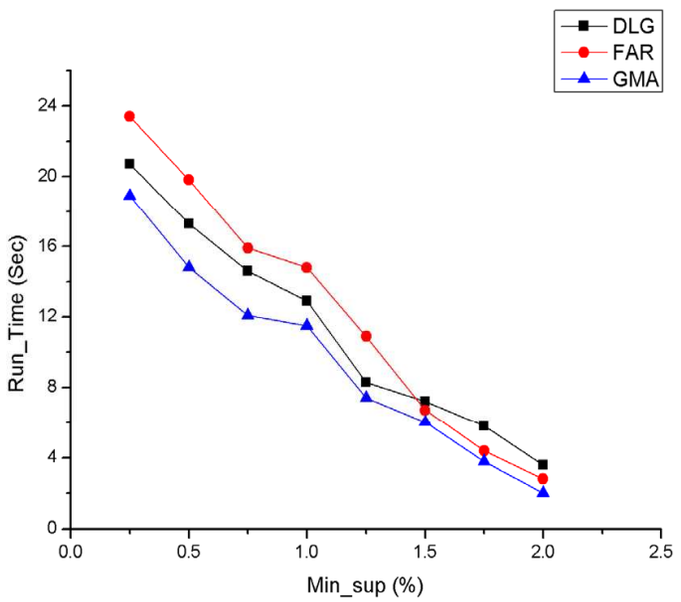


Figure 7. The performance comparisons of DLG,FAR and GMA

Experiment shows that, we can get the same set of frequent-k itemsets using three algorithms. However, the run_time of the algorithms execution is not same. DLG algorithm is better than FAR algorithm when the min_sup is small. FAR algorithm excels DLG algorithm as the value of min_sup increasing. Because when the min_sup increase, the efficiency of matrix pruning is better. In general, GMA algorithm is outperform the two algorithms whatever the value of min_sup is.

5. Further research

Based on the above work, the further research involves the following 3 parts:

i. Semantic cluster concept generation. The above clustering method allows us to automatically obtain the information regarding the spatial layout, the area and the density of a specific cluster. Based on these information, we are able to define a few semantic cluster concepts, such as center cluster, left cluster, dense cluster, sparse cluster, big cluster, small cluster and so on.
ii. Semantic concept image indexing and retrieval. After the generation of cluster semantic concepts, semantic concept indexing of medical images is built to support high-level image retrieval based on these semantic concepts. Examples of such image retrieval are:"retrieval all the medical images which have dense cluster in the center of the image", and "retrieval all the medical images in which the clusters located in the left and lower corners are all small ones".
iii. Trends and patterns mining. Finally, it is desirable to produce some spatial and temporal trends and patterns of the patient who have the medical images with different time. To this end, we explore the pathology cluster information to discover any spatial and temporal trends and patterns of pathology development in terms of scale, area, time duration and location. These trends and patterns are potentially useful for better understanding of the pathology behavior. [3]

6. Conclusion

In this chapter, we firstly presented a progressive water immersion algorithm with guidance of domain knowledge to preprocessing the medical image set to detect objects in medical images. Then we proposed two new algorithms with guidance of domain knowledge to cluster the medical images. We quantified the domain knowledge and use them in the clustering algorithm. Secondly, we proposed GMA algorithm for mining association rules on medical images. GMA algorithm draws on the advantages both association graph and feature matrix pruning to reduce the candidate itemsets generation. In actually, it greatly reduces the candidate itemsets and has improved the efficiency of frequent itemset mining. Experiment shows that GMA algorithm can adapt and adjust better to the change of the value of min_sup. Moreover, GMA algorithm can be used for mining association rules on medical images effectively [9]. We have described the problems with a general form to provide a common framework for other problems appeared in other domains.

Author details

Haiwei Pan
College of Computer Science and Technology,
Harbin Engineering University, Harbin, Heilongjiang, China

Acknowledgement

I would like to express my sincere gratitude to people for helping me during my study. I would especially like to thank my adviser, Jianzhong Li, for his being supportive all the time. I wish to thank Prof. Guisheng Yin, Prof. Jing Zhang, Prof. Qilong Han, Prof. for the advice and the kindness. I would also like to thank Mr. Xiaolei Tan, Ms. Chunxin Zhang, Ms. Guizhen Sun and Mr. Mingde Pan for the research and the support.

The chapter is partly supported by the National Natural Science Foundation of China under Grant No.60803036, No.60803037, Natural Science Foundation of Heilongjiang Province under Grant No.F200903, the Fundamental Research Funds for Central Universities No. HEUCFZ1010, the National High-tech R&D Program of China under Grant No. 2009AA01Z143, and Nuclear Safety & Simulation Tech. Lab Foundation under Grant No.HEUFN0802.

7. References

[1] Haiwei Pan, Jianzhong Li, Wei Zhang(2007) Incorporating domain knowledge into medical image clustering, Applied Mathematics and Computation, 15 February, V185, 844–856

[2] Osmar R. Zaiane, Jiawei Han, Ze-Nian Li, Jean Hou (1998) Mining Multimedia Data CASCON'98: Meeting of Minds, Toronto, Canada, November, 83-96.

[3] WYNNE HSU, MONG LI LEE, JI ZHANG (2002) Image Mining: Trends and Developments. Journal of Intelligent Information Systems, 2002(19-1):7–23.

[4] P. Stolorz, H. Nakamura, E. Mesrobian, R. Muntz, E. Shek, J. Santos, J. Yi, K. Ng, S. Chien, C. Mechoso, and J. Farrara (1995) Fast spatio-temporal data mining of large geophysical datasets. In Proc. Int. Conf. on KDD, 300–305.

[5] U. M. Fayyad, S. G. Djorgovski, and N. Weir (1996) Automating the analysis and cataloging of sky surveys. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 471–493.

[6] O. R. Zaiane, J. Han, Z.-N. Li, J. Y. Chiang, and S. Chee (1998) MultiMediaMiner: A system prototype for multimedia data mining. In Proc. ACM-SIGMOD, Seattle, 581-583.

[7] Osmar R. Zaiane, Jiawei Han, Hua Zhu (2000) Mining Recurrent Items in Multimedia with Progressive Resolution Refinement. Int. Conf. on Data Engineering (ICDE'2000), San Diego, CA, February, 461-470.

[8] Ordonez, C. and Omiecinski, E. (1999). Discovering Association Rules Based on Image Content. In IEEE Advances in Digital Libraries Conference.

[9] Haiwei Pan, Jianzhong Li, Wei Zhang(2005) Medical Image Clustering for Intelligent Decision Support, Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, September, 3308-3311

[10] Wynne Hsu, Mong Li Lee, Kheng Guan Goh (2000) Image Mining in IRIS: Integrated Retinal Information System, In Proc. 2000 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'00). 2000:176~177

[11] Kitamoto.A (2001) Data Mining for Typhoon Image Collection. In Second International Workshop on Multimedia Data Mining (MDM/KDD'2001). 2001:68~78

[12] Ashraf Elsayed, Frans Coenen, Marta García-Fiñana, Vanessa Sluming(2009) Segmentation for Medical Image Mining, A Technical Report, 24 June.

[13] Jain, A.K., Murty, M.N., and Flynn, P.J. (1999) Data Clustering: A Review. ACM Computing Survey, 1999:31~58

[14] Chang, E., Li, C., and Wang, J. (1999) Searching Near-Replicas of Image via Clustering. In SPIE Multimedia Storage and Archiving Systems VI. 1999:89~100

[15] Yu, D. and Zhang, A. (2000) ACQ: An Automatic Clustering and Querying Approach for Large Image Databases. IEEE International Conference on Data Engineering. 2000:58~69

[16] Wynne Hsu, Mong Li Lee, Kheng Guan Goh. (2000) Image Mining in IRIS: Integrated Retinal Information System. Proceedings of the ACM SIGMOD, Dellas, Texas, U.S.A., May 2000, pp. 593.

[17] Maria-Luiza Antonie, Osmar R. Zaiane, Alexandru Coman. (2001) Application of Data Mining Techniques for Medical Image Classification. Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001).

[18] Osmar R. Zaiane, Maria-Luiza Antonie, Alexandru Coman. (2002) Mammography Classification by an Association Rule-based Classifier. Proceedings of the Third International Workshop on Multimedia Data Mining (MDM/KDD'2002).

[19] Christian Bohm, Karin Kailing, Peer Kroger, Arthur Zimek. (2004) Computing Clusters of Correlation Connected Objects. In Proc. 2004 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'04), 2004:455-466.

[20] Anthony K. H. Tung, Xin Xu, Beng Chin Ooi (2005) CURLER: Finding and Visualizing Nonlinear Correlation Clusters, Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), 2005: 518~529

[21] Hans-Peter Kriegel,Martin Pfeifle (2005) Density-Based Clustering of Uncertain Data, Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2005:672~677

[22] Rakesh Agrawal, Tomasz Imielinski and Arun N. Swami (1993) Data Mining: A Performance perspective, IEEE Transactions on Knowledge and Data Engineering, 1993, Vol.5, pp.914-925.

[23] Rakesh Agrawal and Ramakrishnan Srikant (1994) Fast Algorithms for Mining Association Rules in Large Databases, Proceedings of the Twentieth International Conference on Very Large Databases, Santiago, Chile, pp.487-499.

[24] Mika Klemettinen, Heikki Mannila, Pirjo Ronkainen, Hannu Toivonen and A. Inkeri Verkamo (1994) Finding Interesting Rules From Large Sets of Discovered Association

Rules, Proceedings of the Third International Conference on Information and Knowledge Management (CIKM'94), Gaithersburg, USA. pp.401-407.

[25] Jiawei Han, Jian Pei, Yiwen Yin (2000) Mining frequent patterns candidate generation," In proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD'2000), Dallas, TX, 2000, pp.1-12.

[26] J.S. Park, M.S. Chen, P.S. Yu (1995) An effective hash-based algorithm for mining association rules, Proc. 1995 ACM-SIGMOD Int.Conf. Management of Data (SIGMOD'95), San Jose, CA, l995, pp.175-186.

[27] Liu, D. & Kedem, Z. (2002) An Efficient Algorithm for Discovering The Maximum Frequent Set, IEEE Transaction on Knowledge and Data Engineering, 2002, Vol.14, No.3, pp.553-566.

[28] Savasere, E.Ohniecinski, S.Navathe (1995) An efficient algorithm for mining association rules in large databases, Proc.1995 Int. Conf. Very Large Data Bases (VLDB'95), Zurich, Switzerland, l995, pp.432-443.

[29] L.Jaba Sheela , V. Shanthi , D.Jeba Singh (2009) Image Mining using Association rules derived from Feature Matrix, In proceedings of the 2009 International Conference on Advances in Computing, Communication and Control (ICAC3'09), Mumbai, Maharashtra, India, 2009, pp.440~443.

[30] YEN SJ, CHEN ALP (1996) An Efficient Approach to Discovering Knowledge from Large Database, Proceedings of the IEEE/ACM International Conference on Parallel and Distributed Information Systems, Los Angeles, USA, 1996, pp.8~18.

[31] YEN SJ, CHEN ALP (2001) A Graph-Based Approach for Discovering Various Types of Association Rules, IEEE Transactions on Knowledge and Data Engineering, 2001, Vol.13, No.5, pp.839~845.

[32] Pratima Gautam, K. R. Pardasani(2010) A Fast Algorithm for Mining Multilevel Association Rule Based on Boolean Matrix, International Journal on Computer Science and Engineering, Vol. 02, No. 03, 746-752

[33] Haiwei Pan, Xiaolei Tan, Qilong Han, Guisheng Yin(2011) Information Engineering and Electronic Business, A Domain Knowledge Based Approach for Medical Image Retrieval, March, 16-22

[34] Bin Fang, Wynne Hsu, Mong Li Lee(2002) Tumor Cell Identification using Features Rules, SIGKDD , July 23-26