Mining Multimedia Data

SBBD'2000 - Joao Pessoa, Paraiba, Brazil

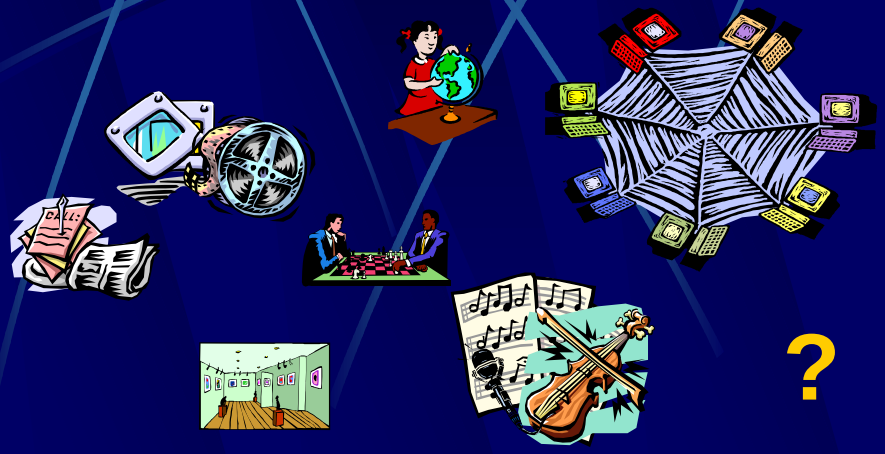
Osmar R. Zaïane

University of Alberta

University of Alberta



Multimedia



?

Outline

- Knowledge Discovery and Data Mining
- Confusion with MDM
- Mining from Video
- Mining from Images
- Mining from Sound

Many Steps in KD Process

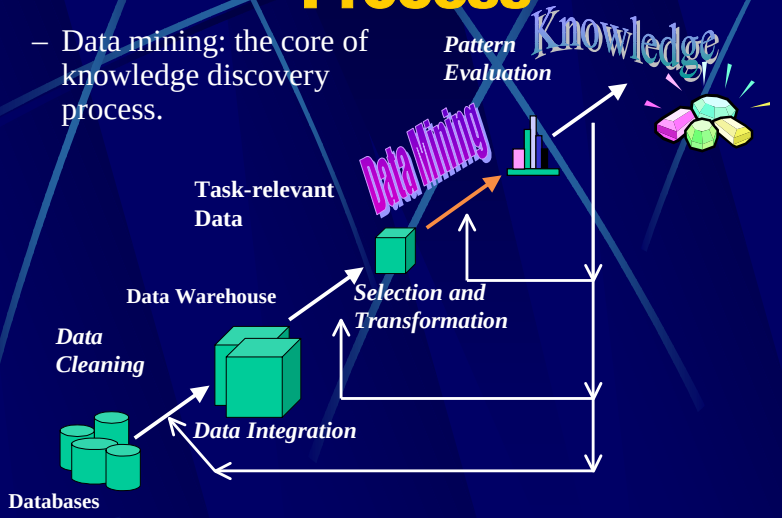
- Gathering the data together
- Cleanse the data and fit it in together
- Select the necessary data
- Crunch and squeeze the data to extract the *essence* of it
- Evaluate the output and use it

So What Is Data Mining?

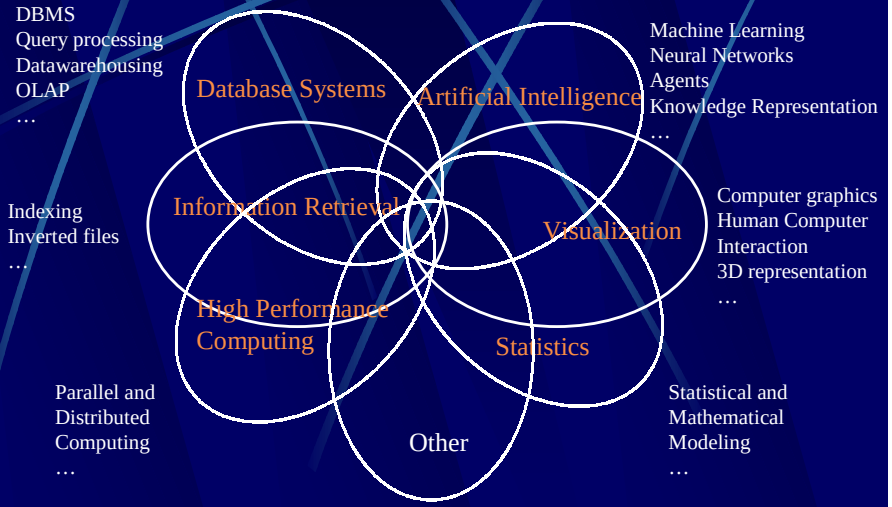
- **In theory, *Data Mining* is a step in the knowledge discovery process. It is the extraction of implicit information from a large dataset.**
- In practice, data mining and knowledge discovery are becoming synonyms.
- There are other equivalent terms: KDD, knowledge extraction, discovery of regularities, patterns discovery, data archeology, data dredging, business intelligence, information harvesting…

Data Mining: A KDD Process

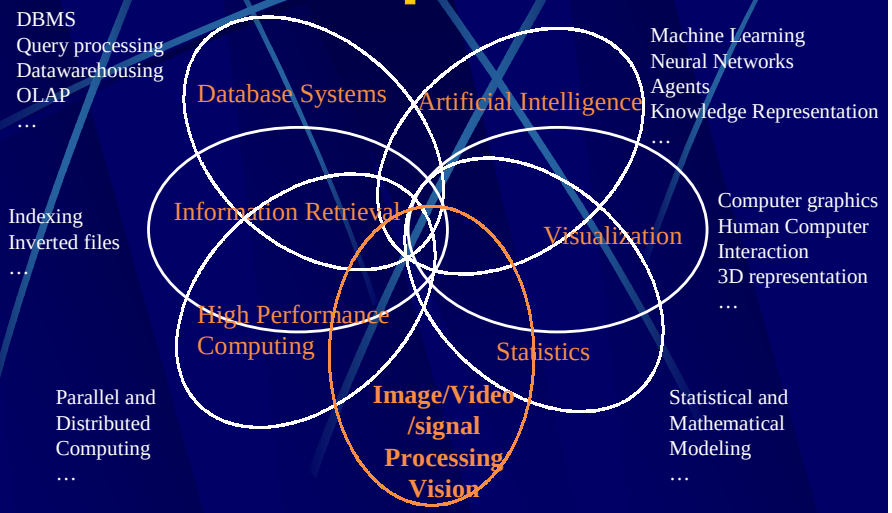
– Data mining: the core of knowledge discovery process.



KDD at the Confluence of Many Disciplines



KDD at the Confluence of Many Disciplines



Outline

- Knowledge Discovery and Data Mining
- **Confusion with MDM**
- Mining from Video
- Mining from Images
- Mining from Sound

Confusion

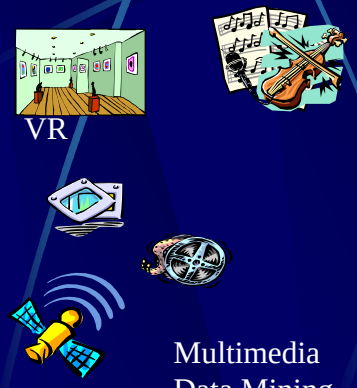


If multimedia subsumes everything, is every data mining multimedia mining?

Are Web mining and multimedia mining the same thing?

No!
Multimedia mining is not mining FROM the Web.
It is better to define or restrict the type of media we consider.

Classes of Mining Problems



Multimedia Data Mining



Web Mining
- Content
- Structure
- Usage

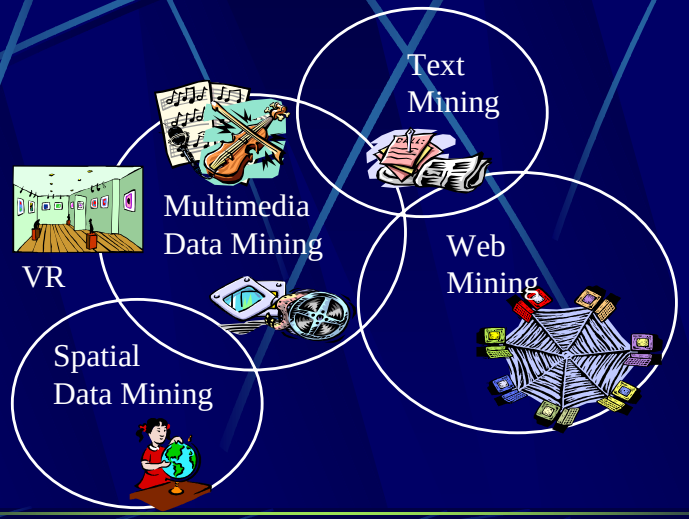


Text Mining
- Unstructured
- Semi-structured

Geo-Spatial Data Mining
- Spatial
- Non-spatial



The Big Picture



Types of MDM



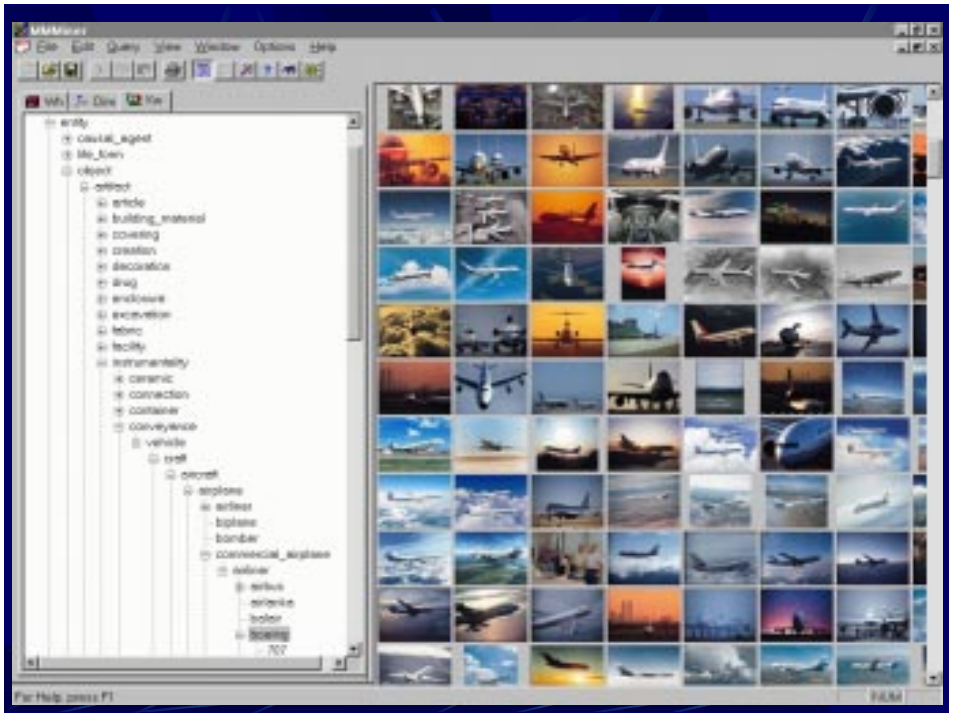
- Use of multimedia in KDD
- Mining multimedia descriptors (metadata)

Not real MDM

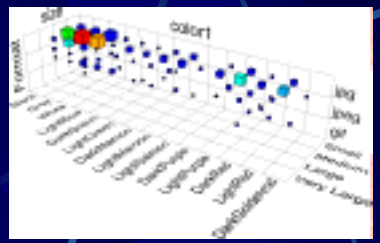
- Extraction of features from multimedia for a higher level application
- Pure multimedia mining

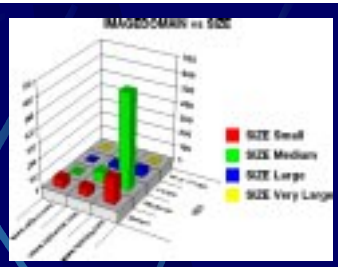
MDM

Content-Based Multimedia Mining vs.
Non Content-Based Multimedia Mining

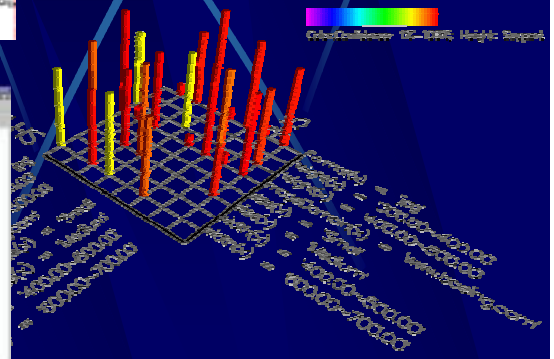


Multimedia OLAP





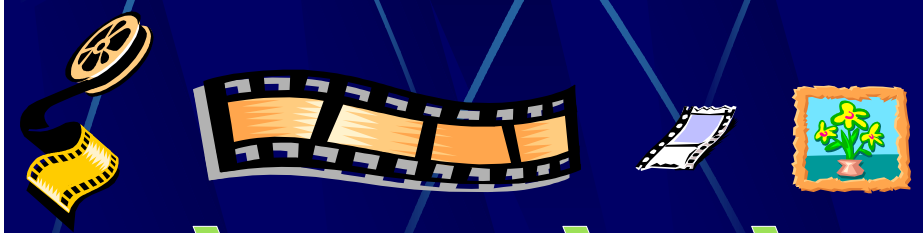




Outline

- Knowledge Discovery and Data Mining
- Confusion with MDM
- **Mining from Video**
- Mining from Images
- Mining from Sound

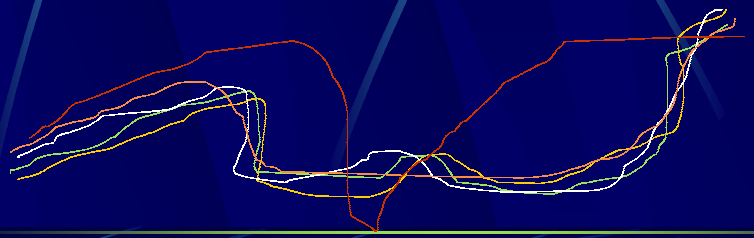
Converting Video to Still Images



Examples of Video Mining



- Mining surveillance camera data
  - Detecting trajectories with stereo cameras
  - Discovering outliers in trajectories
  - Raymond Ng et al. 1998



Examples of Video Mining



- Mining commercial movies
  - Extracting content features such as class of events (violence, happiness…), explosion, rudeness, etc.
  - Expressing sequences of events
  - Correlating with metadata (actors, box office, budget, etc.) to discover rules
  - If ending=s(explosion,violence)➔ W1 > $5M
  - Daniel Barbara et al. 2000

Outline

- Knowledge Discovery and Data Mining
- Confusion with MDM
- Mining from Video
- **Mining from Images**
- Mining from Sound

Image Mining Process

- **Collection and retrieval phase** (Collecting images)
- **Image selection** (Choosing relevant images for the task at hand)
- **Image pre-processing** (Extracting visual features)
- **Mining** (Discovering patterns at individual image level of image group level)
- **Analysis** (validating and interpreting the discovered patterns)

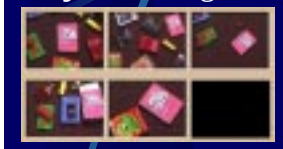
Associations in Medical Images



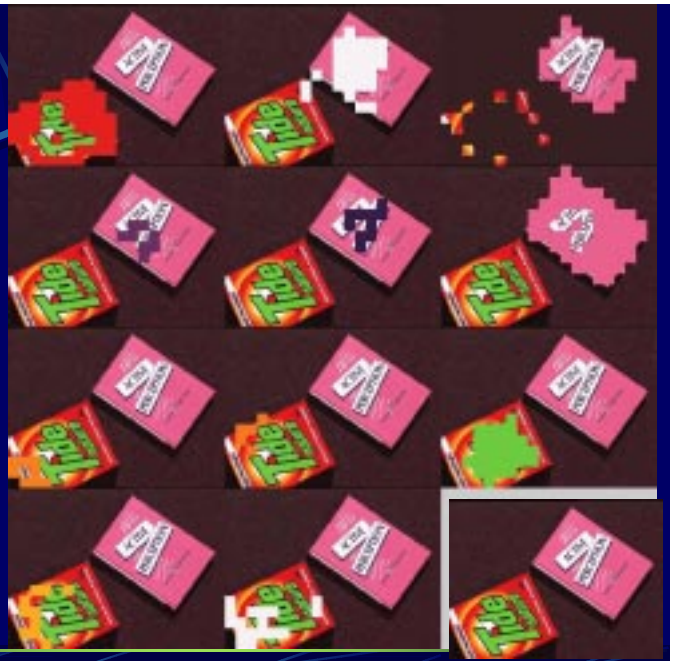
- Associations: presence of lesions, relative positions and spatial relationships.
- Associations with diagnoses and attributes in patient records.

Locales: Colour Localization

➔ Search by **Object** in images



- Locales can have ***any shape;***
- Locales are ***not*** necessarily ***disjoint;***
- Locales can be ***disconnected;***
- The set of locales is ***not*** necessarily ***complete.***

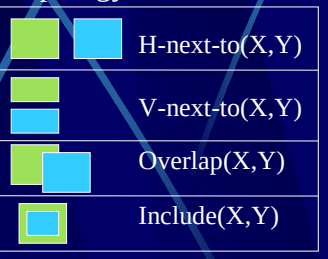


Locales and Their Features

Visual

Colour(X, *colour*)
Size(X, *size*)
Texture(X, *texture*)
Shape(X, *shape*)

Topology

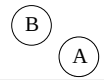
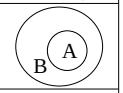
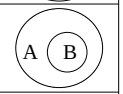
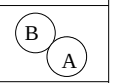
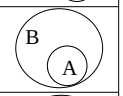
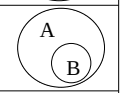


Location

Vertical(X, *v*)
Horizontal(X, *h*)

Movement

Motion(X, *motion*)
Speed(X, *speed*)

$A^\circ \cap B^\circ$	$\delta A \cap \delta B$	$\delta A \cap B^\circ$	$A^\circ \cap \delta B$		
∅	∅	∅	∅	A disjoint B	
¬∅	∅	¬∅	∅	A inside B	
¬∅	∅	∅	¬∅	A contains B	
¬∅	¬∅	∅	∅	A equals B	
∅	¬∅	∅	∅	A meets B	
¬∅	¬∅	¬∅	∅	A covered by B	
¬∅	¬∅	∅	¬∅	A covers B	
¬∅	¬∅	¬∅	¬∅	A overlaps B	

Association Rules

- **Multimedia Association Rule with Recurrent Items:** associate visual object features in images or video frames:

$$\alpha_1 P_1 \wedge \alpha_2 P_2 \wedge \ldots \wedge \alpha_n P_n \rightarrow \lambda_1 Q_1 \wedge \lambda_2 Q_2 \wedge \ldots \wedge \lambda_m Q_m \ (c\%)$$

P_i, $i \in [1..n]$, Q_j, $j \in [1..m]$ are predicates instantiated to topological, visual and kinematics descriptors, α_i and λ_j are integers, αP is true iff P has α occurrences, and c is the confidence.

How to Find Associations

- Find all frequent items (more frequent than minimum support);
- Combine frequent items into itemsets;
- Find frequent itemsets;
- Use frequent itemsets for produce association rules.

➔The problem is to find frequent itemsets
Most famous algorithm is Apriori (R. Agrawal 1994).
There are many other variations and improvements.
However, Apriori misses all item-sets with recurrent items. The re-occurrence of an object can be more significant that its existence in the image.

Transaction-Based vs. Object-Based

Transaction-based support:
Support of an object is the percentage of transactions containing the object. (Number of T_{ob}/ Number of Transactions)

Object-based support:
Support of an object is the percentage of objects equal to the object. (Number of Object occurrences/ Number of Objects)

Finding Itemsets with Reoccurring Items

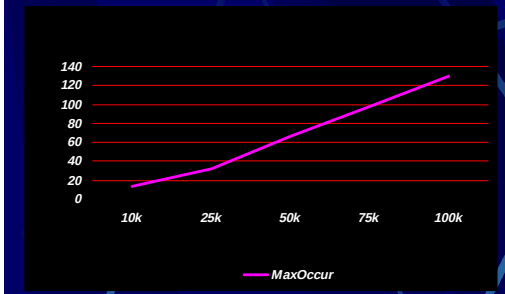
- Devised a new algorithm MaxOccur (zaiane ICDE'2000) (variation of Appriori)
- Finds the maximum time an object can reoccur in an image
- Us max occurrence to increase candidate itemsets during the join

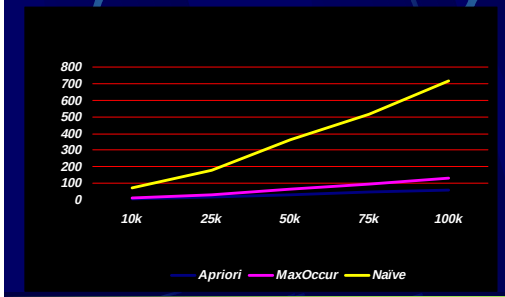
Multimedia Association Rules With Recurrent Spatial Relationships

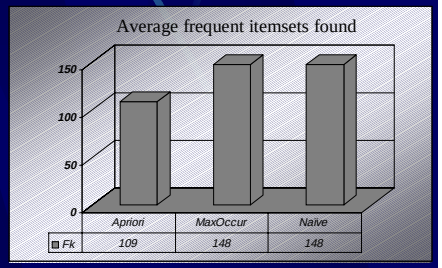
- A spatial relationship is a relationship between 2 objects.
- A spatial relationship is frequent only if the pair of objects is frequent.
- A pair of objects is frequent only if each object is frequent.

Overlap (O_1,O_2) frequent ➔ (O_1 , O_2) frequent ➔ O_1 and O_2 frequent

1. Calculate Frequent Atomic Items;
2. Calculate Frequent Pairs of Atomic Items;
3. Calculate Frequent combinations of Spatial Predicates and Pairs of Atomic Items;
4. Use the set in (3) as Frequent 1-item-sets and Call MaxOccur.







Progressive Resolution Refinement

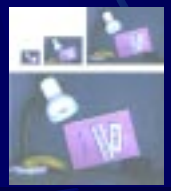
Progressive Resolution Refinement

i=0; D_0 =D;
while (i<maxResLevel) do {
R_i ={sufficiently frequent item-sets at resolution i}
i=i+1; D_i=Filter(D_{i-1}, R_{i-1});
}

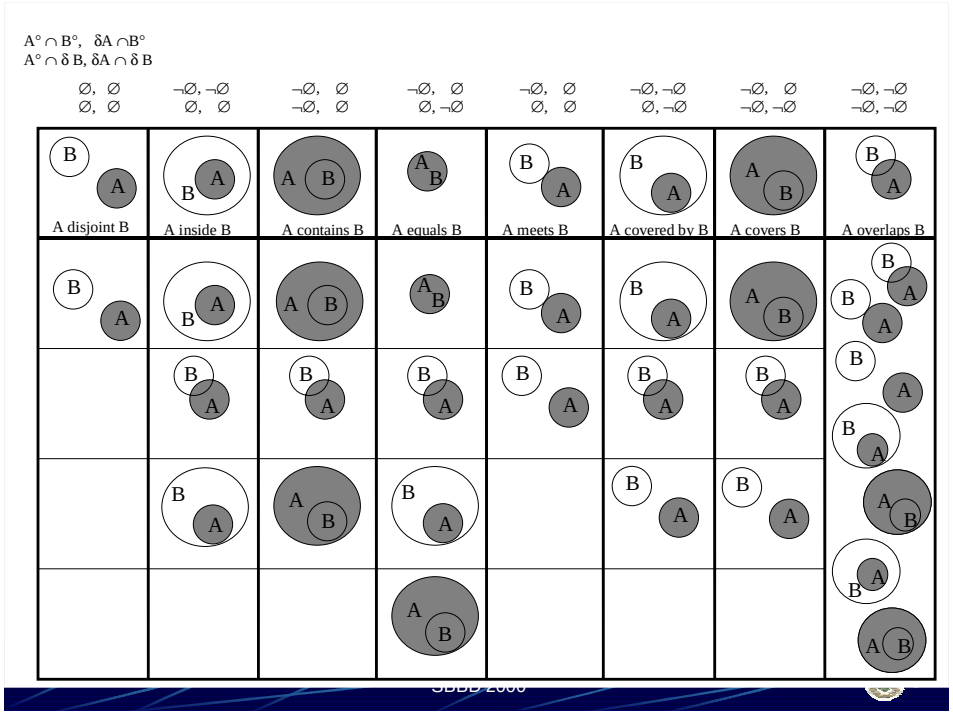
	Feature Localization	Minimum bounding circles	Tile Size
Coarse resolution			
Fine resolution			

From Coarse to Fine Resolution Mining

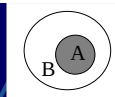
Progressively mine finer resolutions only on candidate frequent item-sets







The topological relation between two areas A and B at any resolution level is defined by a matrix $\mathcal{R}$:



$$\mathcal{R}(A,B)=\begin{bmatrix} A^\circ \cap B^\circ & \delta A \cap B^\circ \\ A^\circ \cap \delta B & \delta A \cap \delta B \end{bmatrix}$$

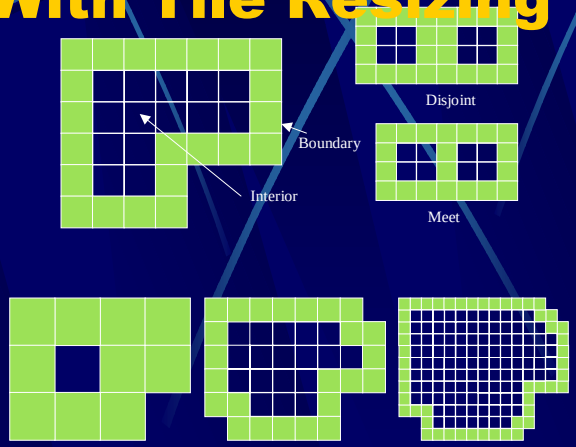
A inside B ➔ $A^\circ \cap B^\circ = \neg\varnothing$ and $\delta A \cap B^\circ = \neg\varnothing$ and
$A^\circ \cap \delta B = \varnothing$ and $\delta A \cap \delta B = \varnothing$

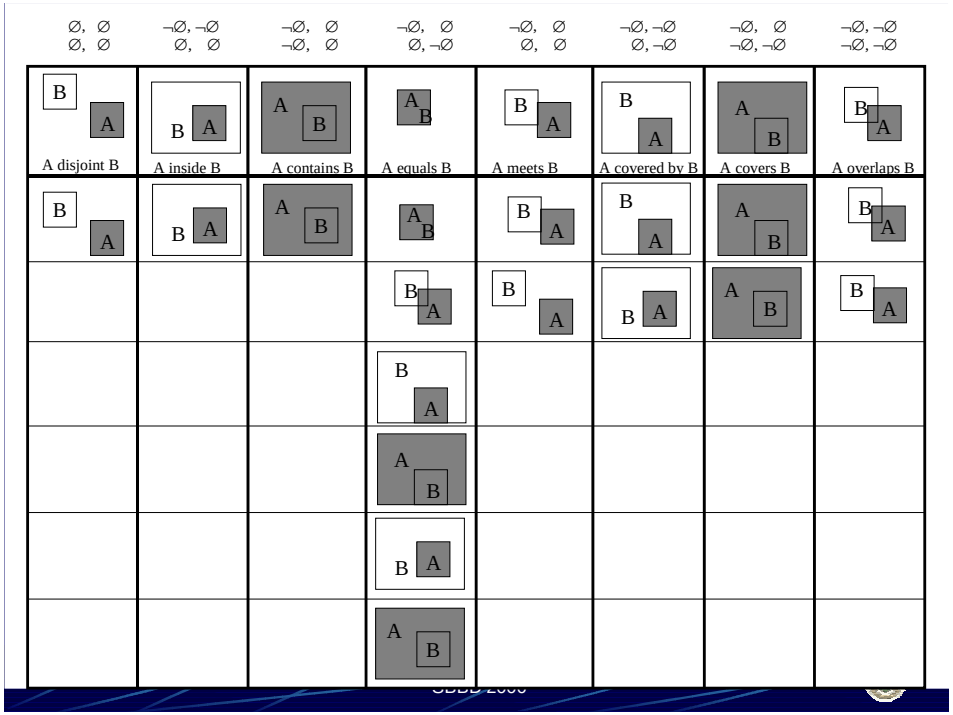
Given ***a*** and ***b*** higher resolutions of **A** and **B**.

$A^\circ \cap B^\circ = \neg\varnothing$ ➔ $a^\circ \cap b^\circ = \neg\varnothing$
$\delta A \cap B^\circ = \neg\varnothing$ ➔ $\delta a \cap b^\circ = \neg\varnothing$
$A^\circ \cap \delta B = \varnothing$ ➔ $a^\circ \cap \delta b = \varnothing$ or $a^\circ \cap \delta b = \neg\varnothing$
$\delta A \cap \delta B = \varnothing$ ➔ $\delta a \cap \delta b = \varnothing$ or $\delta a \cap \delta b = \neg\varnothing$

$$\mathcal{R}(a,b)=\begin{bmatrix} \neg\varnothing & \neg\varnothing \\ \varnothing & \varnothing \end{bmatrix} \text{ or } \begin{bmatrix} \neg\varnothing & \neg\varnothing \\ \neg\varnothing & \neg\varnothing \end{bmatrix} \text{ or } \begin{bmatrix} \neg\varnothing & \neg\varnothing \\ \varnothing & \neg\varnothing \end{bmatrix}$$

Resolution Refinement With Tile Resizing





Outline

- Knowledge Discovery and Data Mining
- Confusion with MDM
- Mining from Video
- Mining from Images
- **Mining from Sound**

State of the Art

- Not much has been done on knowledge discovery from sound
- Classification of high level sound segments using neural networks.
- Mining speech is usually done after conversion speech-to-text or close caption.
- Noting is done on music (as far as I know)

Problem Definition

- Sound Recognition
- Sound Indexing
- Sound Segment Identification
- Noise Filtering
- Compression

Analysis Procedure

- Digitize sound sample
- Segment sound according to some heuristic
  - speech: 1/10th sec. intervals
- Convert sound segments to frequency domain using the FFT
- Classify segments using feature identification derived from FFT analysis

Classification Problem

- How do you classify individual sound segments?
- Using supervised techniques
  - hand tag individual segments
  - use pattern recognition to data set
- What if there are too many segments or too many sound samples?

Unsupervised Classification

- Use of clustering can automate the analysis of sound segment classification
- Needs a robust technique that allows complex similarity metrics to be used
- The ROCK algorithm works well

Clustering Sound with Rock

- We examined:
  - frequency composition
  - harmonic composition
- Devised
  - frequency composition comparison metric
  - harmonic composition comparison metric

Results

- Only speech sounds were analyzed
- ROCK works well but is sensitive during the initial clustering phase to θ threshold settings
- Used modifications to address sensitivity issues
  - Trimming (to reduce “gravity” effect)
  - Threshold Searching

Summary

- Multimedia data mining one phase, in the knowledge discovery process, that extract implicit patterns from large multimedia content.
- Multimedia mining vs. text mining, web mining, spatial mining.
- Usually: video, images, sound.