RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-36

MR. DHIRENDRA

ASSISTANT PROFESSOR

RAMA UNIVERSITY

OUTLINE

- EXAMPLE: ANALYTICAL CHARACTERIZATION
- EXAMPLE: ANALYTICAL COMPARISON
- EXAMPLE2: ANALYTICAL CHARACTERIZATION
- EXAMPLE: ANALYTICAL EXAMPLE: ANALYTICAL COMPARISON
- EXAMPLE: ANALYTICAL EXAMPLE: ANALYTICAL COMPARISON
- MCQ
- REFERENCES

Example: Analytical Characterization

Task

– Compare graduate and undergraduate undergraduate students using students using

discriminant rule.

– DMQL query

use Big_University_DB

mine comparison as “grad_vs_undergrad_students”

in relevance to in relevance to name gender major birth place birth date residence name, gender, major, birth_place, birth_date, residence, phone#, gpa

for “graduate_students”

where status in “graduate”

versus “undergraduate_students”

where status in “undergraduate”

analyze count%

from std t

Example: Analytical comparison

Given

– attributes attributes name, gender, major, birth place birth_place,

birth_date, residence, phone# and gpa

– Gen(a) = concept hierarchies on attributes a i) = concept hierarchies on attributes ai

– Ui = attribute analytical thresholds for attributes ai

– T = attribute generalization thresholds for i = attribute generalization thresholds for

attributes ai

– R = attribute attribute relevance relevance threshold threshold

Example2: Analytical Characterization

1. Data collection

– target and contrasting classes

• 2. Attribute Attribute relevance relevance analysis analysis

– remove attributes name, gender, major, phone#

• 3. Synchronous generalization

– controlled by user-specified dimension thresholds

– prime target and contrasting contrasting classes classes relations/cuboids relations/cuboids

Example: Analytical Example: Analytical comparison

Birth_country	Age_range	Gpa	Count%
Canada	20-25	Good	5.53%
Canada	25-30	Good	2.32%
Canada	Over_30	Very_good	5.86%
…	…	…	…
Other	Over_30	Excellent	4.68%

Prime generalized relation for the target class: Graduate students

Birth_country	Age_range	Gpa	Count%
Canada	15-20	Fair	5.53%
Canada	15-20	Good	4.53%
…	…	…	…
Canada	25-30	Good	5.02%
…	…	…	…
Other	Over_30	Excellent	0.68%

Prime generalized relation for the contrasting class: Undergraduate students

Example: Analytical Example: Analytical comparison

4. Drill down, roll up and other OLAP operations operations on target and contrasting classes to adjust levels of abstractions of resulting description

• 5. Presentation

– as generalized relations, crosstabs, bar charts, pie charts, or rules

– contrasting measures to reflect comparison between target and contrasting contrasting classes classes

• count%

Multiple Choice Question

1. Various visualization techniques are used in _____ step of KDD.
a) selection
b) transformaion
c) data mining.
d) interpretation.

2. Extreme values that occur infrequently are called as _____.
a) outliers
b) rare values.
c) dimensionality reduction.
d) All of the above.

3. Box plot and scatter diagram techniques are _____.
a) Graphical
b) Geometric
c) Icon-based.
d) Pixel-based.

4. _____ is used to proceed from very specific knowledge to more general information.
a) Induction
b) Compression.
c) Approximation.
d) Substitution.

5. Describing some characteristics of a set of data by a general model is viewed as _____
a) Induction
b) Compression
c) Approximation
d) Summarization

REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm
- https://www.geeksforgeeks.org/
- http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf DATA MINING BOOK WRITTEN BY Micheline Kamber
- https://www.javatpoint.com/three-tier-data-warehouse-architecture
- M.H. Dunham, " Data Mining: Introductory & Advanced Topics" Pearson Education
- Jiawei Han, Micheline Kamber, " Data Mining Concepts & Techniques" Elsevier
- Sam Anahory, Denniss Murray," data warehousing in the Real World: A Practical Guide for Building Decision Support Systems, " Pearson Education
- Mallach," Data Warehousing System", TMH
- R. Agrawal, A. Gupta, and S. Sarawagi. Modeling multidimensional databases. ICDE'97 S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. ACM SIGMOD Record, 26:65-74, 1997
- S. Agarwal, R. Agrawal, P. M. Deshpande, A. Gupta, J. F. Naughton, R. Ramakrishnan, and S. Sarawagi. On the computation of multidimensional aggregates. VLDB'96 D. Agrawal, A. E. Abbadi, A. Singh, and T. Yurek. Efficient view maintenance in data warehouses. SIGMOD'97
- E. F. Codd, S. B. Codd, and C. T. Salley. Beyond decision support. Computer World, 27, July 1993.
- J. Gray, et al. Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. Data Mining and Knowledge Discovery, 1:29-54, 1997.