CloseViz: Visualizing Useful Patterns

Chris Carmichael   Carson K. Leung
Department of Computer Science
The University of Manitoba, Canada

UP @ KDD 2010

Outline

- Introduction
- Motivation & related work
  - Existing visualizers
- Proposed visualizer
  - *CloseViz*: Visualizing closed frequent patterns
- Conclusions

Carmichael & Leung (U Manitoba, Canada)

Introduction & Motivation

- Focus on the KDD of *frequent pattern mining*

- <u>Motivation</u>: Since the introduction of frequent pattern mining, lots of algorithms have been developed
    - They mostly return the mined results in *textual forms*
- "A picture is worth a thousand words"
    - ***Visual representation*** helps users in gaining insight into massive amounts of data or information

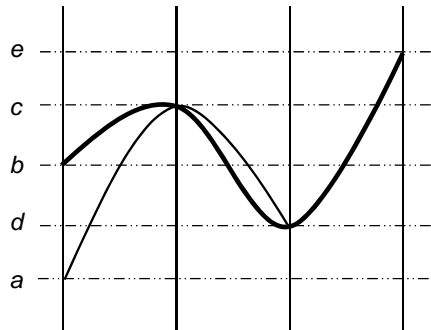Carmichael & Leung (U Manitoba, Canada)

Motivation: Existing Visualizers

- Many were designed to visualize *association rules*

    (e.g., {apples, bananas} ➔ {cherries, dates})

- Recently, there are visualizers that can be used for visualizing *frequent patterns*

Carmichael & Leung (U Manitoba, Canada)
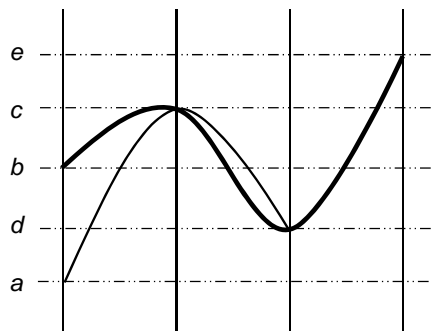
A Sample Visualizer #1

- Designed to visualize *association rules*
- Can be used for visualizing *frequent patterns*

- Uses a 2D space consisting of many vertical axes
- Evenly distributes domain items along these vertical axes
- Represents an itemset X as a curve
- Uses thickness of the curve to indicate frequency of an itemset X

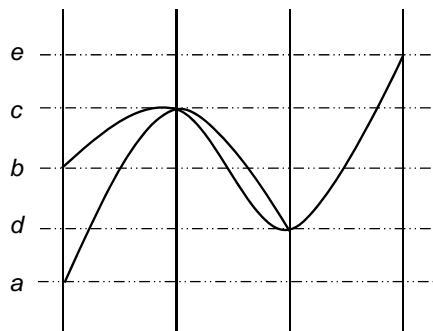Carmichael & Leung (U Manitoba, Canada)

A Sample Visualizer #1

- {*a,c,d*}, {*b,c,d,e*}
- frequency(*e*) ≥ frequency(*c*) ≥ frequency(*b*) ≥ frequency(*d*) ≥ frequency(*a*)

Carmichael & Leung (U Manitoba, Canada)

A Sample Visualizer #1

- Do these curves represent itemsets…
    - $\{a,c,d\}$ & $\{b,c,d,e\}$
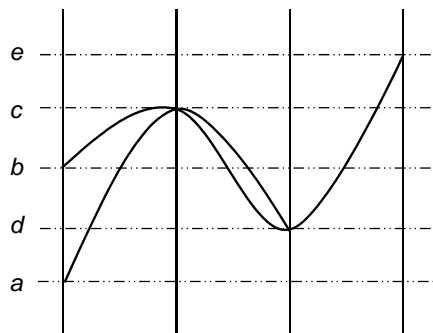 - or
    - $\{a,c,d,e\}$ & $\{b,c,d\}$?

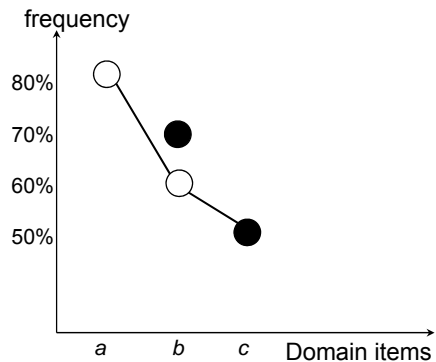# A Sample Visualizer #1

Problems:

1. Does not clearly show the (absolute) frequency of an domain item
2. Not easy to tell the (absolute) frequency of an itemset by judging the thickness of curves
3. Curves cross over each other

A Sample Visualizer #2: FIsViz [PAKDD'08]

- Designed to visualize *frequent patterns*

- Uses a 2D space with domain items on the x-axis & frequency on the y-axis
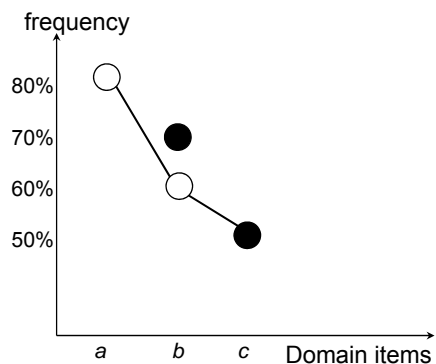- Represents an itemset X as a polyline

Carmichael & Leung (U Manitoba, Canada)

A Sample Visualizer #2: FIsViz

Advantages:
1. Clearly shows the frequency of an domain item
   - E.g., frequency({b}) = 70%
2. Easy to tell the frequency of an itemset
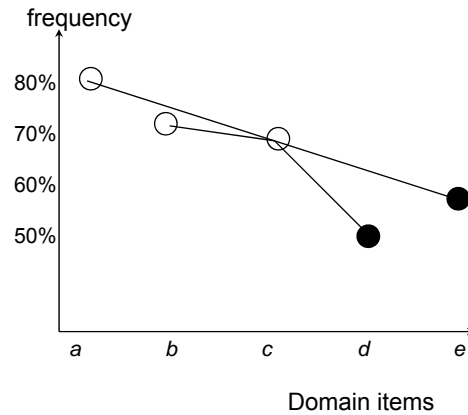   - E.g., freq({a,b,c}) = 50%

Carmichael & Leung (U Manitoba, Canada)
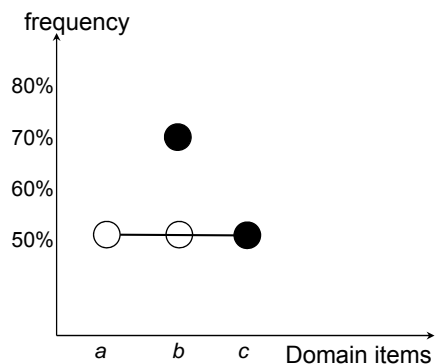
A Sample Visualizer #2: FIsViz

Potential problem:
- Polylines bend & cross over each other
- E.g., do these polylines represent itemsets…

  {*a,c,d*} & {*b,c,e*}

  or

  {*a,c,e*} & {*b,c,d*}?

frequency

80%
70%
60%
50%

a   *b*   *c*   *d*   *e*

Domain items

Carmichael & Leung (U Manitoba, Canada)

A Sample Visualizer #3: WiFIsViz [ICDM'08]

- Also designed to visualize *frequent patterns*

- Uses a 2D space with domain items on the x-axis & frequency on the y-axis
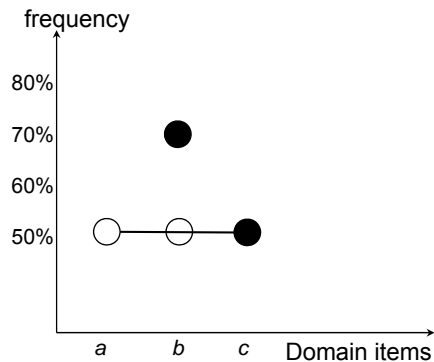- Represents an itemset X as a **horizontal line**

frequency

80%
70%
60%
50%

a   *b*   *c*   Domain items

Carmichael & Leung (U Manitoba, Canada)

A Sample Visualizer #3: WiFIsViz

Advantages:

1. Clearly shows the frequency of an domain item
   - E.g., frequency({*b*}) = 70%
2. Easy to tell the frequency of an itemset
   - E.g., freq({*a,b,c*}) = 50%

Carmichael & Leung (U Manitoba, Canada)

A Sample Visualizer #3: WiFIsViz

Potential problems:

1. Shows **all** frequent patterns
   - Lots of horizontal lines
2. Multiple frequent patterns may have the same frequency
   - Broad band for each frequency value
     or
   - Many horizontal lines project onto one ➔ info loss ({a,b,c,d} is at 60% or 50%?)
3. Uses different icons (unfilled vs. filled circles)
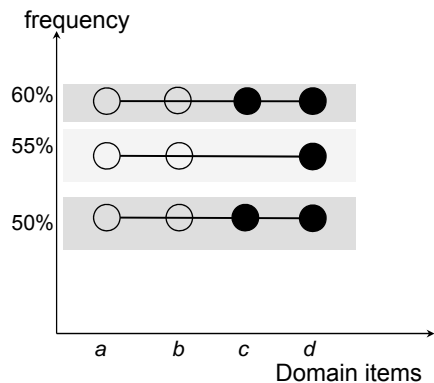
Carmichael & Leung (U Manitoba, Canada)

Problem Statement

- We provide users with **a visualizer that is designed for showing only useful patterns** & that avoids aforementioned potential problems

- <u>Contribution</u>: We propose CloseViz (which shows closed frequent patterns)

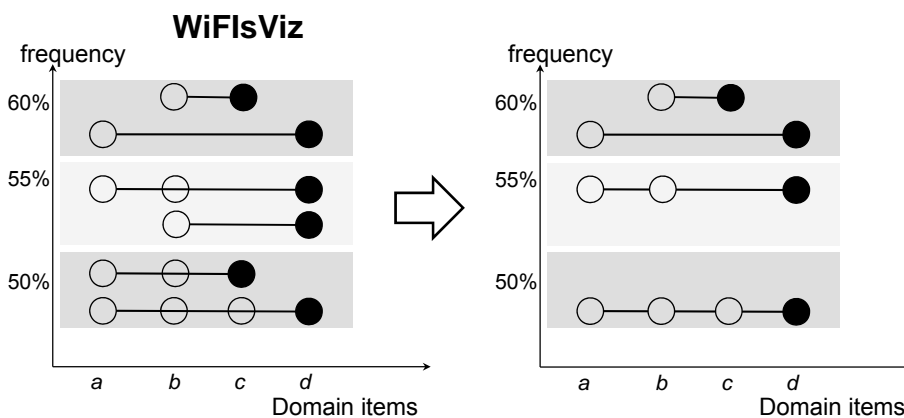Carmichael & Leung (U Manitoba, Canada)

Our Visualizer: CloseViz

CloseViz

- Like WiFIsViz, CloseViz ...
  - uses a 2D space with domain items on the x-axis & frequency on the y-axis
  - represents an itemset X as a **horizontal line**
- Unlike WiFIsViz, CloseViz ...
  - shows closed frequent patterns (instead of all frequent patterns)
  - uses only one type of icons (i.e., unfilled circle)
  - distinguishes real patterns vs. the results of projection

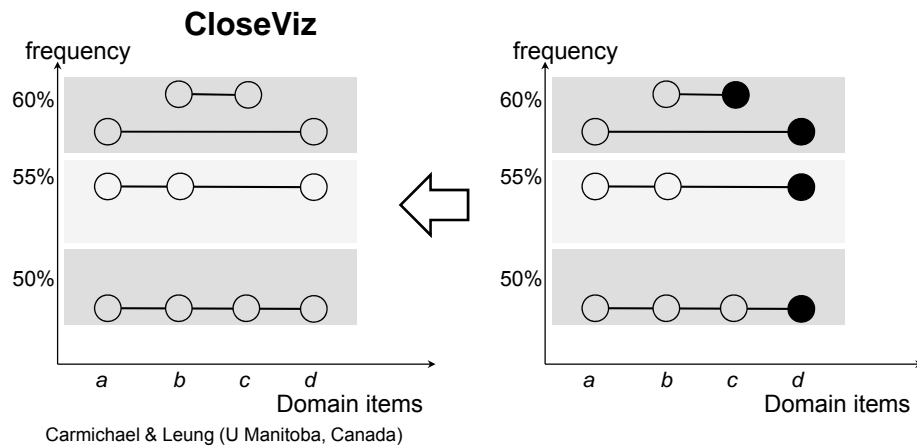Carmichael & Leung (U Manitoba, Canada)

CloseViz

1. Shows closed frequent patterns
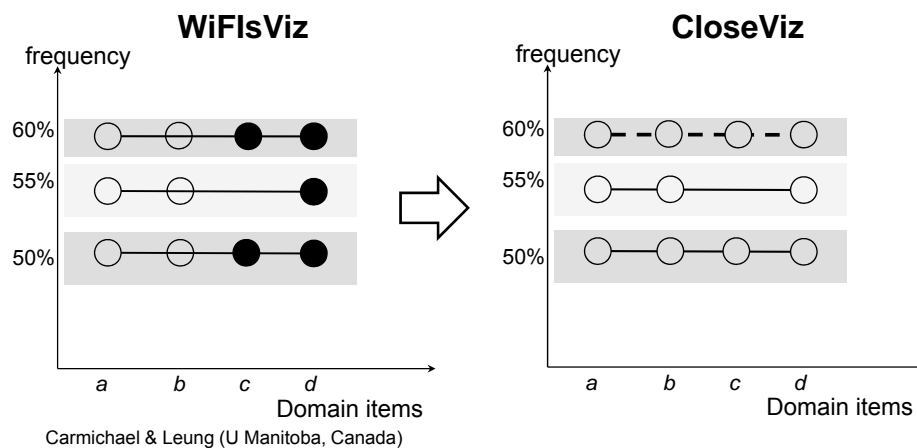

Carmichael & Leung (U Manitoba, Canada)
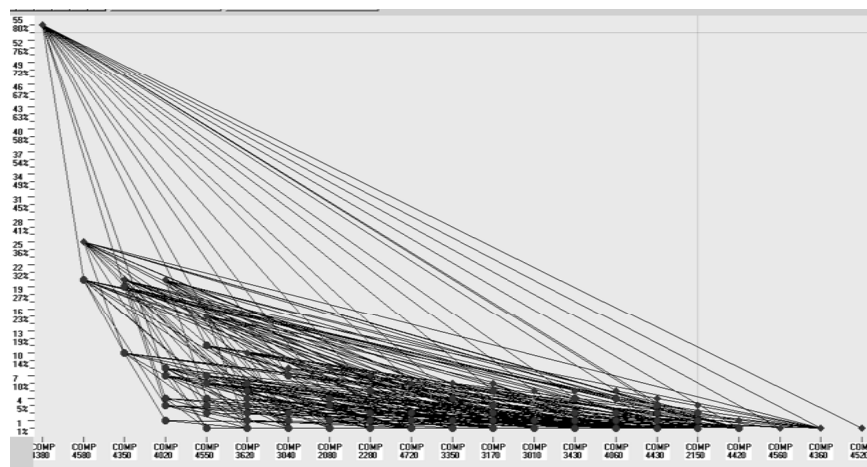
CloseViz

2. Uses only unfilled circles

CloseViz

Carmichael & Leung (U Manitoba, Canada)

CloseViz

3. Represents real closed patterns by *solid lines*, results of projection by *dashed lines*

WiFIsViz   **CloseViz**

Carmichael & Leung (U Manitoba, Canada)
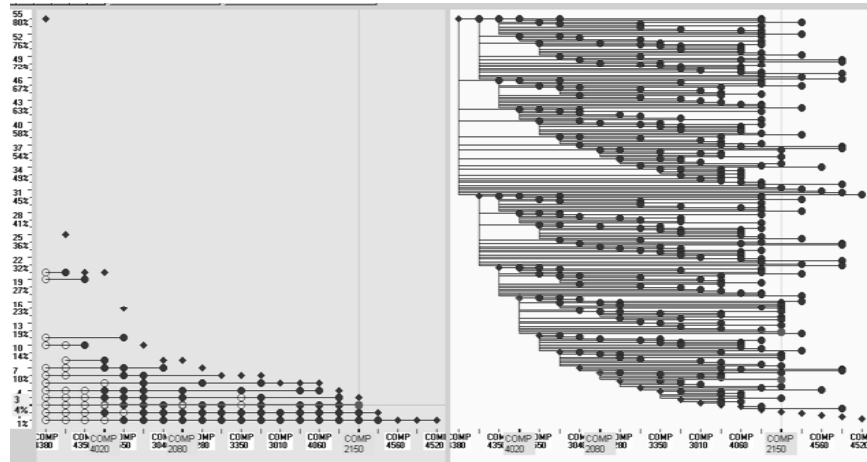
Sample Screenshots

Carmichael & Leung (U Manitoba, Canada)
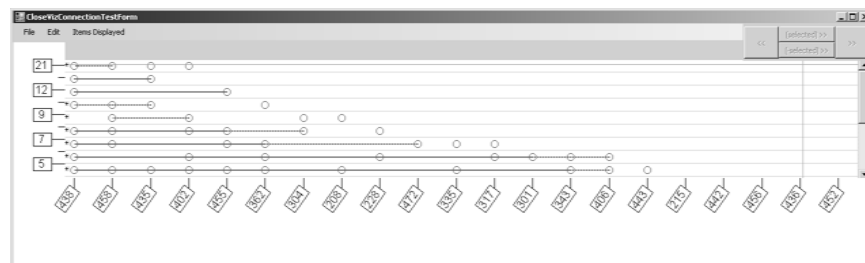
Screenshot of FIsViz

Carmichael & Leung (U Manitoba, Canada)

Screenshot of WiFIsViz

Carmichael & Leung (U Manitoba, Canada)

Screenshot of CloseViz

Carmichael & Leung (U Manitoba, Canada)

Conclusions

Conclusions

- We proposed **CloseViz**, which provides users with a visualizer that ...
  - is designed for showing useful patterns (namely, closed frequent patterns) &
  - avoids aforementioned potential problems of existing visualizers
- CloseViz …
  - reduces #patterns to be shown
  - allows visual exploration
  - retains all important info (➔ Closed patterns can be served as surrogates for all frequent patterns)

Carmichael & Leung (U Manitoba, Canada)

Thank you / Merci

kleung [AT] cs.umanitoba.ca
www.cs.umanitoba.ca/~kleung
dblab.cs.umanitoba.ca