A Connection Between Cluster Analysis and Formal Concept Analysis

Melvin F. Janowitz

DIMACS, Rutgers University
Piscataway, NJ 08854

BCS-CSNA Meeting on February 2, 2007

Background From Jardine and Sibson

Book: Mathematical Taxonomy by N. Jardine and R. Sibson, Wiley, New York, 1971. This got me started.

Background From Jardine and Sibson

Book: Mathematical Taxonomy by N. Jardine and R. Sibson, Wiley, New York, 1971. This got me started.

Underlying finite set to be classified: E

$\Sigma(E)$ the reflexive symmetric relations on E.

Dissimilarity coefficient (DC) $d : E \times E \longrightarrow \Re_0^+$

- $d(a, b) = d(b, a)$
- $d(a, a) = 0$

Background From Jardine and Sibson

Book: Mathematical Taxonomy by N. Jardine and R. Sibson, Wiley, New York, 1971. This got me started.

Underlying finite set to be classified: E

$\Sigma(E)$ the reflexive symmetric relations on E.

Dissimilarity coefficient (DC) $d : E \times E \longrightarrow \Re_0^+$

- $d(a, b) = d(b, a)$
- $d(a, a) = 0$

d is an ultrametric if also

- $d(a, b) \leq \max\{d(a, c), d(b, c)\}$ for all $a, b, c \in E$.

Background From Jardine and Sibson

Book: Mathematical Taxonomy by N. Jardine and R. Sibson, Wiley, New York, 1971. This got me started.

Underlying finite set to be classified: E

$\Sigma(E)$ the reflexive symmetric relations on E.

Dissimilarity coefficient (DC) $d : E \times E \longrightarrow \Re_0^+$

- $d(a,b) = d(b,a)$
- $d(a,a) = 0$

d is an ultrametric if also

- $d(a,b) \leq \max\{d(a,c), d(b,c)\}$ for all $a,b,c \in E$.

Numerically stratified clustering (NSC) $Td : \Re_0^+ \longrightarrow \Sigma(E)$ a residual mapping in that

- There is an h such that $Td(h) = E \times E$.
- $Td(\bigwedge h_i) = \bigcap Td(h_i)$.

NSCs and DCs are in one-one correspondence.

In the book a cluster method is viewed as a transformation of a DC to an ultrametric.

An order theoretic approach

All this suggests an order theoretic framework. This was introduced in a 1978 paper

An order theoretic model for cluster analysis,
SIAM Journal of Applied Math. **34**, 55-72.

But there is a problem.

An order theoretic approach

All this suggests an order theoretic framework. This was introduced in a 1978 paper

An order theoretic model for cluster analysis,
SIAM Journal of Applied Math. **34**, 55-72.

But there is a problem.

- Can measure dissimilarities in a lattice L with smallest element 0.

An order theoretic approach

All this suggests an order theoretic framework. This was introduced in a 1978 paper

An order theoretic model for cluster analysis,
SIAM Journal of Applied Math. **34**, 55-72.

But there is a problem.

- Can measure dissimilarities in a lattice L with smallest element 0.
- Thus a DC becomes $d : E \times E \to L$.

An order theoretic approach

All this suggests an order theoretic framework. This was introduced in a 1978 paper

An order theoretic model for cluster analysis,
SIAM Journal of Applied Math. **34**, 55-72.

But there is a problem.

- Can measure dissimilarities in a lattice L with smallest element 0.
- Thus a DC becomes $d : E \times E \to L$.
- An ultrametric satisfies $d(a, b) \leq d(a, c) \vee d(b, c)$ for all $a, b, c \in E$.

An order theoretic approach

All this suggests an order theoretic framework. This was introduced in a 1978 paper

An order theoretic model for cluster analysis,
SIAM Journal of Applied Math. **34**, 55-72.

But there is a problem.

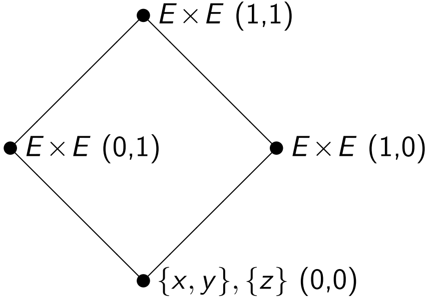
- Can measure dissimilarities in a lattice L with smallest element 0.
- Thus a DC becomes $d : E \times E \to L$.
- An ultrametric satisfies $d(a, b) \leq d(a, c) \vee d(b, c)$ for all $a, b, c \in E$.
- An NSC remains a residual mapping. This is a well known lattice theoretic mapping and is related to what are called Galois connections.

An Example

Example: $E = \{x, y, z\}$ illustrating the problem.
Dissimilarities measured in 2^2. Here is the input DC

Pairs	Attribute 1	Attribute 2
xy	0	0
xz	1	0
yz	0	1

Cluster Method: Single Linkage Clustering



The levels for cluster $E \times E$ do not occur at a smallest level!

A Modified Model

L is poset with smallest element 0 where dissimilarities measured.

$\mathcal{F}(L) =$ order filters of L ordered by

$F \leq G \iff G \subseteq F$.

$F \neq \emptyset$, $x \in F, x \leq y$ impiles $y \in F$.

Principal filter: $F_h = \{y \in L : y \geq h\}$.

$\mathcal{F}(L)$ is a complete distributive lattice.

A Modified Model

L is poset with smallest element 0 where dissimilarities measured.

$\mathcal{F}(L) =$ order filters of L ordered by

$F \leq G \iff G \subseteq F$.

$F \neq \emptyset$, $x \in F, x \leq y$ impiles $y \in F$.

Principal filter: $F_h = \{y \in L : y \geq h\}$.

$\mathcal{F}(L)$ is a complete distributive lattice.

DC: $D : E \times E \to \mathcal{F}(L)$ such that

$D(a, b) = D(b, a)$

$D(a, a) \leq D(a, b)$. Might want $D(a, a) = F_0$.

A Modified Model

L is poset with smallest element 0 where dissimilarities measured.

$\mathcal{F}(L) =$ order filters of L ordered by

$F \leq G \iff G \subseteq F$.

$F \neq \emptyset$, $x \in F, x \leq y$ impiles $y \in F$.

Principal filter: $F_h = \{y \in L : y \geq h\}$.

$\mathcal{F}(L)$ is a complete distributive lattice.

DC: $D : E \times E \to \mathcal{F}(L)$ such that

$D(a, b) = D(b, a)$

$D(a, a) \leq D(a, b)$. Might want $D(a, a) = F_0$.

Can take $D(a, b)$ to be principal filters. Ultrametric if also

$D(a, b) \leq D(a, c) \vee D(b, c)$ for all $a, b, c \in E$.

A Modified Model

L is poset with smallest element 0 where dissimilarities measured.

$\mathcal{F}(L)$ = order filters of L ordered by

$F \leq G \iff G \subseteq F$.

$F \neq \emptyset$, $x \in F, x \leq y$ impiles $y \in F$.

Principal filter: $F_h = \{y \in L : y \geq h\}$.

$\mathcal{F}(L)$ is a complete distributive lattice.

DC: $D : E \times E \to \mathcal{F}(L)$ such that

$D(a, b) = D(b, a)$

$D(a, a) \leq D(a, b)$. Might want $D(a, a) = F_0$.

Can take $D(a, b)$ to be principal filters. Ultrametric if also

$D(a, b) \leq D(a, c) \vee D(b, c)$ for all $a, b, c \in E$.

$SD : L \to \Sigma(E)$ (Symmetric relations on E).

Gives cluster candidates at level h.

$SD(h) = \{(a, b) : h \in D(a, b)\}$.

$h \leq k$ implies $SD(h) \subseteq SD(k)$.

If L has a largest member 1, then $SD(1) = E \times E$.

$h \in D(a, b) \iff (a, b) \in SD(h)$.

Boolean Dissimilarities

Assume E has k binary attributes. We want to construct $D : E \times E \to 2^k$, where the ith component of $D(x, y)$ depends only on the values $x(i)$, $y(i)$ of the ith attribute for x and y, If $x(i) \neq y(i)$, we want $D(i) = 1$, so we need only worry about $x(i) = y(i)$.

Boolean Dissimilarities

Assume E has k binary attributes. We want to construct $D : E \times E \longrightarrow 2^k$, where the ith component of $D(x,y)$ depends only on the values $x(i)$, $y(i)$ of the ith attribute for x and y, If $x(i) \neq y(i)$, we want $D(i) = 1$, so we need only worry about $x(i) = y(i)$. For $x \neq y$, there are only three possibilities:

(D_1) $D(i) = 0$ if $x(i) = y(i) = 1$ and 1 otherwise

(D_2) $D(i) = 0$ if $x(i) = y(i) = 0$ and 1 otherwise

(D_3) $D(i) = 1$ if $x(i) \neq y(i)$ and 0 otherwise.

Boolean Dissimilarities

Assume E has k binary attributes. We want to construct $D : E \times E \to 2^k$, where the ith component of $D(x,y)$ depends only on the values $x(i)$, $y(i)$ of the ith attribute for x and y, If $x(i) \neq y(i)$, we want $D(i) = 1$, so we need only worry about $x(i) = y(i)$. For $x \neq y$, there are only three possibilities:

(D_1) $D(i) = 0$ if $x(i) = y(i) = 1$ and 1 otherwise

(D_2) $D(i) = 0$ if $x(i) = y(i) = 0$ and 1 otherwise

(D_3) $D(i) = 1$ if $x(i) \neq y(i)$ and 0 otherwise.

For $x = y$, we could either take $D(i) = 0$, or $\bigvee_y \{D(x,y) : x \neq y\}$.

Fact: Each D_i is an ultrametric.

Boolean Dissimilarity Example with $E = \{x, y, z\}$

Objects	A1	A2	A3
x	1	1	0
y	1	0	1
z	0	1	0

D_1	x	y	z
x	001	011	101
y	011	010	111
z	111	111	101

Boolean Dissimilarity Example with $E = \{x, y, z\}$

Objects	A1	A2	A3
x	1	1	0
y	1	0	1
z	0	1	0

D_1	x	y	z
x	001	011	101
y	011	010	111
z	111	111	101

D_2	x	y	z
x	110	111	110
y	111	101	111
z	110	111	010

D_3	x	y	z
x	000	011	100
y	011	000	111
z	100	111	000

Boolean Dissimilarity Example with $E = \{x, y, z\}$

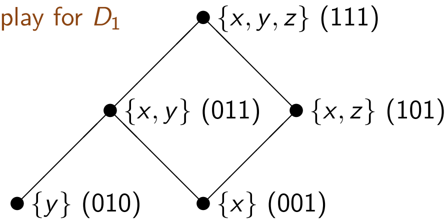
Objects	A1	A2	A3
x	1	1	0
y	1	0	1
z	0	1	0

D_1	x	y	z
x	001	011	101
y	011	010	111
z	111	111	101

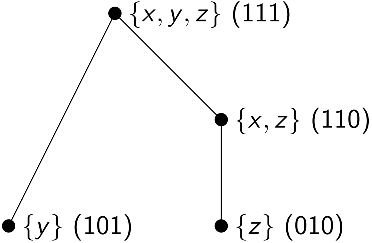
D_2	x	y	z
x	110	111	110
y	111	101	111
z	110	111	010

D_3	x	y	z
x	000	011	100
y	011	000	111
z	100	111	000

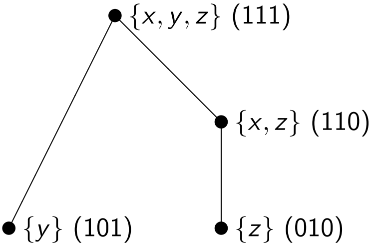
Display for D_1



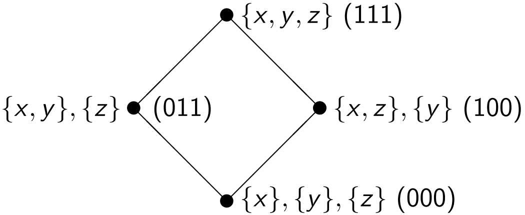
Display for D_2



Display for D_2



Display for D_3



Formal Concept Analysis – thumbnail sketch

Formal Context $\mathfrak{G}, \mathfrak{M}$ sets with

$\perp \subseteq \mathfrak{G} \times \mathfrak{M}$ binary relation called the incidence relation.

Members of $\mathfrak{G}$ objects, $\mathfrak{M}$ attributes.

$A^{\perp} = \{m \in \mathfrak{M} : a \perp m \ \forall a \in A\}$. $B^{\perp} = \{g \in \mathfrak{G} : b \perp g \ \forall b \in B\}$.

$(A \subseteq \mathfrak{G}, B \subseteq \mathfrak{M})$

Formal Concept Analysis – thumbnail sketch

Formal Context $\mathfrak{G}, \mathfrak{M}$ sets with

$\perp \subseteq \mathfrak{G} \times \mathfrak{M}$ binary relation called the incidence relation.

Members of $\mathfrak{G}$ objects, $\mathfrak{M}$ attributes.

$A^\perp = \{m \in \mathfrak{M} : a \perp m \ \forall a \in A\}$. $B^\perp = \{g \in \mathfrak{G} : b \perp g \ \forall b \in B\}$.

$$(A \subseteq \mathfrak{G}, B \subseteq \mathfrak{M})$$

(A, B) formal concept if $A = B^\perp, B = A^\perp$.

Extent is A. *Intent* is B.

For concepts $(A, B), (C, D)$,

$A \subseteq C$ iff $D \subseteq B$.

Order the concepts by inclusion of extents.

Get a complete lattice.

Formal Concept Analysis – thumbnail sketch

Formal Context $\mathfrak{G}, \mathfrak{M}$ sets with

$\perp \subseteq \mathfrak{G} \times \mathfrak{M}$ binary relation called the incidence relation.

Members of $\mathfrak{G}$ objects, $\mathfrak{M}$ attributes.

$A^\perp = \{m \in \mathfrak{M} : a \perp m \; \forall a \in A\}$. $B^\perp = \{g \in \mathfrak{G} : b \perp g \; \forall b \in B\}$.

$(A \subseteq \mathfrak{G}, B \subseteq \mathfrak{M})$

(A, B) formal concept if $A = B^\perp, B = A^\perp$.

Extent is A. *Intent* is B.

For concepts $(A, B), (C, D)$,

$A \subseteq C$ iff $D \subseteq B$.

Order the concepts by inclusion of extents.

Get a complete lattice.

Nice reference: Introduction to Lattices and Order by Davey and Priestley.

The formal concept approach

Objects	A1	A2	A3
x	1	1	0
y	1	0	1
z	0	1	0

$\mathfrak{G} = \{x, y, z\}$ $\qquad$ $\mathfrak{M} = \{A1, A2, A3\}$

The formal concept approach

Objects	A1	A2	A3
x	1	1	0
y	1	0	1
z	0	1	0

$\mathfrak{G} = \{x, y, z\}$ $\quad\quad$ $\mathfrak{M} = \{A1, A2, A3\}$

$A1^{\perp} = \{x, y\}$ $A2^{\perp} = \{x, z\}$ $A3^{\perp} = \{y\}$

The formal concept approach

Objects	A1	A2	A3
x	1	1	0
y	1	0	1
z	0	1	0

$\mathfrak{G} = \{x, y, z\} \qquad \mathfrak{M} = \{A1, A2, A3\}$

$A1^{\perp} = \{x, y\}$ $A2^{\perp} = \{x, z\}$ $A3^{\perp} = \{y\}$

$$\{x, y\}^{\perp} = \{A1, A2\} \cap \{A1, A3\} = \{A1\}$$
$$\{x, z\}^{\perp} = \{A1, A2\} \cap \{A2\} = \{A2\}$$
$$\{y\}^{\perp} = \{A1, A3\}.$$

The formal concept approach

Objects	A1	A2	A3
x	1	1	0
y	1	0	1
z	0	1	0

$\mathfrak{G} = \{x, y, z\}$ $\quad$ $\mathfrak{M} = \{A1, A2, A3\}$

$A1^{\perp} = \{x, y\}$ $A2^{\perp} = \{x, z\}$ $A3^{\perp} = \{y\}$

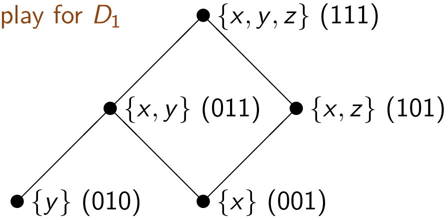
$\{x, y\}^{\perp} = \{A1, A2\} \cap \{A1, A3\} = \{A1\}$
$\{x, z\}^{\perp} = \{A1, A2\} \cap \{A2\} = \{A2\}$
$\{y\}^{\perp} = \{A1, A3\}$.

List of formal concepts:

$(\{x, y, z\}, \emptyset)$ $\quad$ $(\{x, y\}, A1)$ $\quad$ $(\{x, z\}, A2)$
$(\{y\}, \{A1, A3\})$ $\quad$ $(\{x\}, \{A1, A2\})$

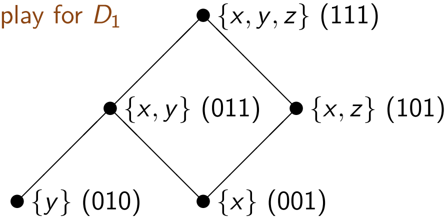
Comparison of the two approaches

Display for D_1

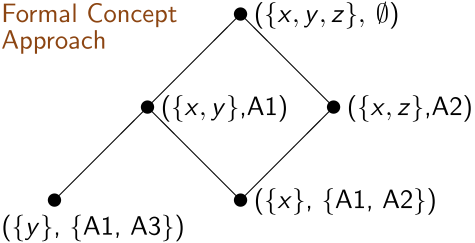


Comparison of the two approaches

Display for D_1



Formal Concept Approach



Complete-Linkage Algorithm

$d : E \times E \rightarrow P$ a poset.

Complete-linkage algorithm:
Assume P is image of d.

1. For each minimal element m, form transitive closure of $Td(m)$. This is output at level m. Call it $F(m)$.
2. Look at levels h that cover minimal members m

Complete-Linkage Algorithm

$d : E \times E \to P$ a poset.

Complete-linkage algorithm:
Assume P is image of d.

1. For each minimal element m, form transitive closure of $Td(m)$. This is output at level m. Call it $F(m)$.
2. Look at levels h that cover minimal members m
3. For each such h, form $\cup\{F(m) : m < h, m \text{ minimal }\}$. These are all clusters at level h. Add any pairs from $T(h)$. Use complete linkage criterion to merge any clusters.
4. continue the process.

An example involving numerical data

		Water	Protein	Fat	Lactose	Ash
1.	Bison	86.9	4.8	1.7	5.7	0.9
2.	Buffalo	82.1	5.9	7.9	4.7	0.78
3.	Camel	87.7	3.5	3.4	4.8	0.71
4.	Cat	81.6	10.1	6.3	4.4	0.75
5.	Deer	65.9	10.4	19.7	2.6	1.4
6.	Dog	76.3	9.3	9.5	3.0	1.2

Composition of Mammal Milk

An example involving numerical data

		Water	Protein	Fat	Lactose	Ash
1.	Bison	86.9	4.8	1.7	5.7	0.9
2.	Buffalo	82.1	5.9	7.9	4.7	0.78
3.	Camel	87.7	3.5	3.4	4.8	0.71
4.	Cat	81.6	10.1	6.3	4.4	0.75
5.	Deer	65.9	10.4	19.7	2.6	1.4
6.	Dog	76.3	9.3	9.5	3.0	1.2

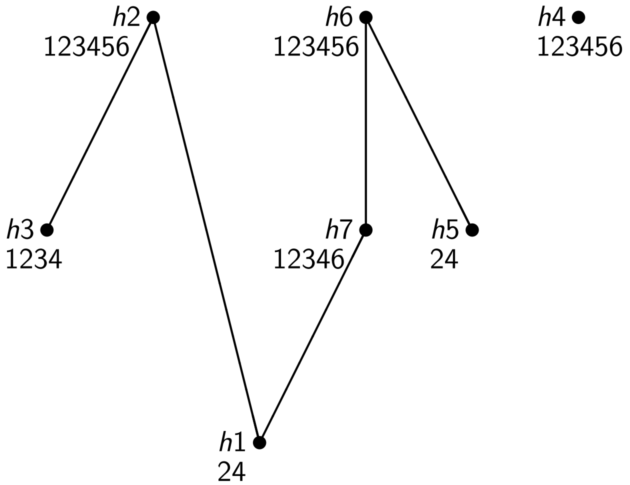
Composition of Mammal Milk

Original data has 25 mammal species. Just wanted a short example. Used dissimilarity taking values in $\mathcal{Z}^5$ where $\mathcal{Z}$ denotes positive integers. Here is construction. Used squared Euclidean distance on each attribute to construct five separate dissimilarities, then represented them as columns in a single dissimilarity matrix having 15 rows and 5 columns. We use the vector ordering inherited from $\mathcal{Z}^5$. To simplify notation, we just rank ordered each column of the matrix. This then produced a dissimilarity taking values in a subposet of $\mathcal{Z}^5$, as displayed in the following table.

Label	object pair	Water	Protein	Fat	Lactose	Ash
$m1$	12	3	3	8	5	4
$m2$	13	2	5	2	4	6
$h1$	14	4	11	6	6	5
$h2$	15	13	12	14	13	12
$h3$	16	9	9	9	12	8
$m3$	23	5	6	5	1	3
$m4$	24	1	8	1	2	1
$h4$	25	12	10	11	10	13
$m5$	26	6	7	1	8	9
$h5$	34	7	14	3	3	2
$h6$	35	14	15	13	11	15
$h7$	36	10	13	7	9	11
$m6$	45	11	1	12	9	14
$m7$	46	4	2	4	7	10
$m8$	56	8	4	10	3	7

Non-minimal Clusters of the Output

For purposes of comparison, standard complete linkage clustering yields the proper clusters 13, 24, 246, 12346.


h2
123456
h6
123456
h4
123456
h3
1234
h7
12346
h5
24
h1
24