Pattern Structures for Identifying Biclusters with Coherent Sign Changes

Nyoman Juniarta   Miguel Couceiro   Amedeo Napoli

Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

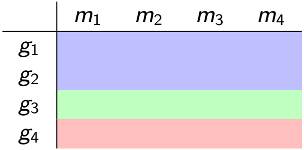
SFC 2019

Outline

Biclustering

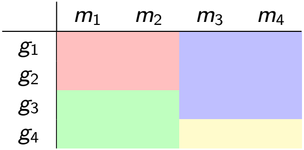
Clustering

Find groups of similar objects.



Biclustering

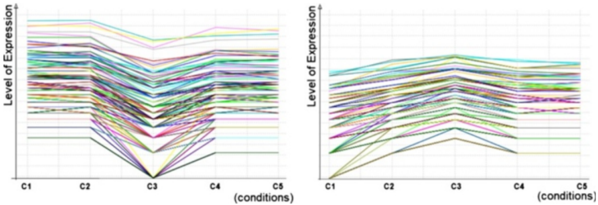
Find groups of similar objects w.r.t. a set of attributes.



Biclustering

An application: order preserving biclustering[1]

Genes	Conditions			
	c_1	c_2	c_3	c_4
g_1	1	2	5	2
g_2	5	7	8	9
g_3	1	2	3	2
g_4	2	1	3	9



[1] Henriques, R., & Madeira, S. C. (2014). BicSPAM: Flexible biclustering using sequential patterns. *BMC Bioinformatics*, 15(1), 130.

Coherent-sign-changes

Searching a set of experimental conditions which affect a set of genes in a consistent way.

Genes	Conditions				
	c_1	c_2	c_3	c_4	c_5
g_1	$+$	$+$	$+$	$-$	$+$
g_2	$-$	$+$	$+$	$-$	$+$
g_3	$-$	$-$	$-$	$+$	$-$
g_4	$-$	$+$	$+$	$-$	$+$
g_5	$-$	$-$	$-$	$-$	$+$

: any pair of conditions are either identical or opposite $\Rightarrow$ coherent-sign-changes (CSC) bicluster.

Outline

Formal concept analysis (FCA)

	m_1	m_2	m_3	m_4
g_1	×			×
g_2	×	×	×	×
g_3			×	×
g_4	×	×		×

Formal context $= (G, M, I)$
$I \subseteq G \times M$
ex. $(g_2, m_4) \in I$

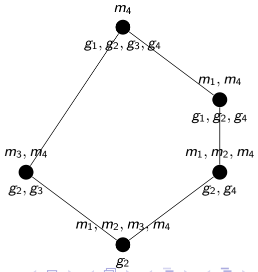
$A \subseteq G, B \subseteq M$
$A' = \{m \in M | \forall g \in A, (g, m) \in I\}$
$B' = \{g \in G | \forall m \in B, (g, m) \in I\}$

ex.
$\{g_2, g_4\}' = \{m_1, m_2, m_4\}$
$\{m_1, m_2, m_4\}' = \{g_2, g_4\}$

A formal concept is a pair (A, B) where $A' = B$ and $B' = A$

$(\{g_2, g_4\}, \{m_1, m_2, m_4\})$ is a formal concept

	m_1	m_2	m_4	m_3
g_1	×		×	
g_2	×	×	×	×
g_4	×	×	×	
g_3			×	×



Pattern structures

FCA

Formal context $= (G, M, I)$
G is a set of objects
M is a set of attributes
$I \subseteq G \times M$

$A \subseteq G, B \subseteq M$
$A' = \{m \in M | \forall g \in A, (g, m) \in I\}$
$B' = \{g \in G | \forall m \in B, (g, m) \in I\}$

A formal concept is a pair (A, B) where
$A' = B$ and $B' = A$
A is extent and B is intent

Pattern Structures

Pattern structures $= (G, (D, \sqcap), \delta)$
G is a set of objects
$(D, \sqcap)$ is a lattice of descriptions
$\delta : G \to D$

$A \subseteq G, d \in D$
$A^\diamond = \prod_{g \in \mathtt{A}} \delta(g)$
$d^\diamond = \{g \in G | d \sqsubseteq \delta(g)\}$

A pattern concept is a pair (A, d) where
$A^\diamond = d$ and $d^\diamond = A$
A is extent and d is intent

Outline

Partition

abc

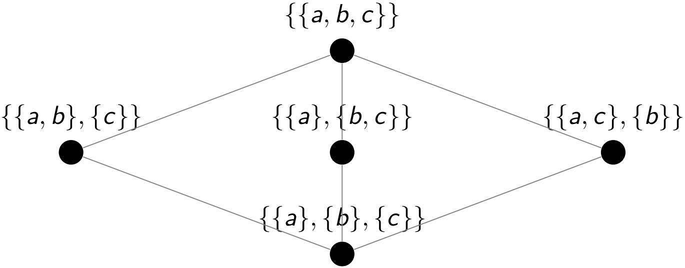
Partition

$$abc$$

$$\overbrace{\{\underbrace{\{a, b\}}_{component}, \{\underbrace{c}_{element}\}\}}^{partition}$$

Partition

The set of partitions is partially ordered.



Partition pattern structures

Pattern Structures

Pattern structures $= (G, (D, \sqcap), \delta)$
G is a set of objects
$(D, \sqcap)$ is a lattice of descriptions
$\delta : G \rightarrow D$

$A \subseteq G, d \in D$
$A^{\diamond} = \bigsqcap_{g \in A} \delta(g)$
$d^{\diamond} = \{g \in G | d \sqsubseteq \delta(g)\}$

A concept is a pair (A, d) where $A^{\diamond} = d$ and $d^{\diamond} = A$
A is extent and d is intent

Partition Pattern Structures (PPS)

PPS $= (G, (D, \sqcap), \delta)$
G is a set of objects
$(D, \sqcap)$ is a lattice of attribute partitions
$\delta : G \rightarrow D$

$A \subseteq G, d \in D$
$A^{\diamond} = \bigsqcap_{g \in A} \delta(g)$
$d^{\diamond} = \{g \in G | d \sqsubseteq \delta(g)\}$

A concept is a pair (A, d) where $A^{\diamond} = d$ and $d^{\diamond} = A$
A is extent and d is intent

Partition pattern structures

Definition

Similarity of two partitions $d_1 = \{p_i\}$ and $d_2 = \{p_j\}$: common "coarsest refinement" of partitions

$$d_1 \sqcap d_2 = \bigcup_{i,j} \{p_i \cap p_j\}$$

Example

	m_1	m_2	m_3	m_4
g_1	+	+	−	−
g_2	+	+	−	−
g_3	−	−	+	−
g_4	+	+	+	+
g_5	−	−	−	−

$$\delta(g_2) = \{ \{m_1, m_2\}, \{m_3, m_4\} \}$$

$$\delta(g_3) = \{ \{m_1, m_2, m_4\}, \{m_3\} \}$$

$\cap$ $= \{m_1, m_2\}$

$\cap$ $= \emptyset$

$\cap$ $= \{m_4\}$

$\cap$ $= \{m_3\}$

$$\delta(g_2) \sqcap \delta(g_3) = \{\{m_1, m_2\}, \{m_4\}, \{m_3\}\}$$

Outline

Signed partition

$$\overbrace{\{\underbrace{\{a, b, c\}}_{component}, \{\underbrace{d}_{element}, e\}\}}^{partition}$$

$$\overbrace{\{\underbrace{\{a^+, b^-, c^+\}}_{signed\_component}, \{\underbrace{d^-}_{signed\_element}, e^-\}\}}^{signed\_partition}$$

Description

Definition

Description d of an object g ($\delta(g)$): A partition having one component. This component contains all attributes and their sign.

Example

	m_1	m_2	m_3	m_4
g_1	$+$	$+$	$-$	$-$
g_2	$+$	$+$	$-$	$-$
g_3	$-$	$-$	$+$	$-$
g_4	$+$	$+$	$+$	$+$
g_5	$-$	$-$	$-$	$-$

$$\delta(g_1) = \delta(g_2) = \{\{m_1^+, m_2^+, m_3^-, m_4^-\}\}$$
$$\delta(g_3) = \{\{m_1^-, m_2^-, m_3^+, m_4^-\}\}$$
$$\delta(g_4) = \delta(g_5) = \{\{m_1^+, m_2^+, m_3^+, m_4^+\}\}$$

Signed component equality

First possibility

Two components with *same elements* having *same sign* are equal

$\{a^-, b^-\} = \{a^-, b^-\}$
$\{a^-, b^-\} \neq \{a^+, b^+\}$
$\{a^-, b^-\} \neq \{a^+, b^-\}$

Second possibility

Two components with *same elements* having *opposite sign* are equal

$\{a^-, b^-\} \neq \{a^-, b^-\}$
$\{a^-, b^-\} = \{a^+, b^+\}$
$\{a^-, b^-\} \neq \{a^+, b^-\}$

Our definition

Two components are equal if their elements have *entirely same* or *entirely opposite* sign.

$$\{a^-, b^-\} = \{a^-, b^-\}$$
$$\{a^-, b^-\} = \{a^+, b^+\}$$
$$\{a^-, b^-\} \neq \{a^+, b^-\}$$

Similarity

Definition

$\cap^{\pm}$: Similarity between two signed_components
$\Rightarrow$ Two signed_components: 1) for the same sign, 2) for the different sign.
Similarity of two signed_partitions: $\cap^{\pm}$ between any pair of signed_components.

Example

$d_1 = \{ \{m_1^+, m_2^+, m_4^-\}, \{m_3^+, m_5^-, m_6^-\} \}$

$d_2 = \{ \{m_1^+, m_2^+\}, \{m_3^+, m_4^+, m_5^-, m_6^+\} \}$

	Same	Different
(blue) $\cap^{\pm}$ (red)	$\{m_1^+, m_2^+\}$	$\emptyset$
(blue) $\cap^{\pm}$ (yellow)	$\emptyset$	$\{m_4^-\}$
(green) $\cap^{\pm}$ (red)	$\emptyset$	$\emptyset$
(green) $\cap^{\pm}$ (yellow)	$\{m_3^+, m_5^-\}$	$\{m_6^-\}$

$d_1 \sqcap d_2 = \{\{m_1^+, m_2^+\}, \{m_4^-\}, \{m_3^+, m_5^-\}, \{m_6^-\}\}$

Signed partition pattern structures

Definition

Triple $(G, (D, \sqcap), \delta)$

$$A^{\Box} = \bigsqcap_{g \in A} \delta(g), \qquad A \subseteq G$$

$$d^{\Box} = \{g \in G | d \sqsubseteq \delta(g)\}, \qquad d \in D$$

(A, d) is a concept if $A^{\Box} = d$ and $d^{\Box} = A$.

Example

	m_1	m_2	m_3	m_4
g_1	$+$	$+$	$-$	$-$
g_2	$+$	$+$	$-$	$-$
g_3	$-$	$-$	$+$	$-$
g_4	$+$	$+$	$+$	$+$
g_5	$-$	$-$	$-$	$-$

A concept:
$(\{g_1, g_2, g_3\}, \{\{m_1^+, m_2^+, m_3^-\}, \{m_4^-\}\})$

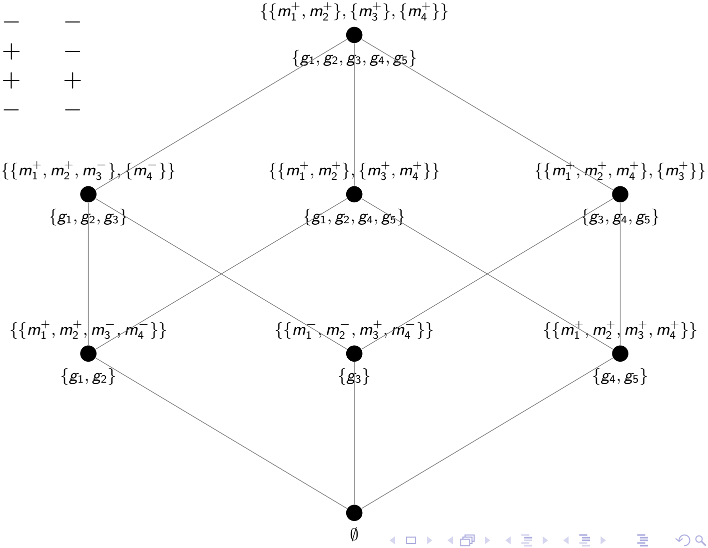
Two CSC biclusters:

$(\{g_1, g_2, g_3\}, \{m_1, m_2, m_3\})$

$(\{g_1, g_2, g_3\}, \{m_4\})$

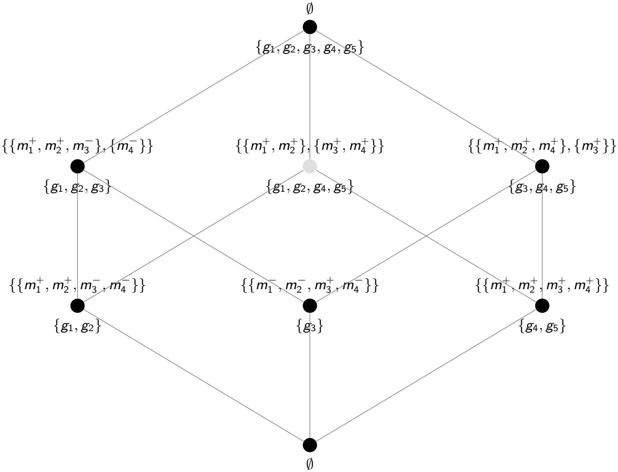
Signed partition pattern structures

	m_1	m_2	m_3	m_4
g_1	+	+	−	−
g_2	+	+	−	−
g_3	−	−	+	−
g_4	+	+	+	+
g_5	−	−	−	−



Small experiment

AddIntent, with parameter θ: min. size of a component that an intent should have. With $\theta = 3$:



Small experiment

Lymphoma dataset: 4026 genes/objects, 96 conditions/attributes

θ	Runtime (minutes)	Number of concepts All	Extent size > 1	Largest extent size
70	229.4	157K	153K	8
80	62.9	7K	2K	7
90	62.1	4K	83	3

Outline

Conclusion

An approach to mine biclusters with coherent sign changes is presented, based on partition pattern structures.
By introducing the notion of *signed_partition*, and formulating the corresponding similarity operator.

Questions/remarks?