Closed patterns and abstraction beyond lattices

but not so far ....

Henry Soldano

LIPN, Université Paris 13, Sorbonne Paris Cité, France

ICFCA - June 2014

Plan

Plan

Recent work on closed patterns in data mining

Closed patterns w.r.t. a set of objects in $F \subseteq 2^X$ can be obtained using a closure operator on F even when F has a structure weaker than a lattice
(Boley and Coll., TCS 2010).

Definition

$F \subseteq 2^P$ is a confluent family iff

- $\emptyset \in F$, and
- for all $t, t_1, t_2 \in F$ with $\emptyset \neq t, t_1 \supseteq t, t_2 \supseteq t$, we have:

$$t_1 \cup t_2 \in F$$

The set of subsets of edges inducing *connected* subgraphs of some graph G is a confluent family. : if an edge x belongs both to the connected subgraphs induced by X_1 and X_2 then the subgraph induced by $X_1 \cup X_2$ is also connected.

Extending FCA using *local* meet operators

We investigate partial orders equipped with a local meet operator, we call pre-confluences.

1. A nice result on sets of closed elements in a pre-confluence
2. Existence closed elements w.r.t a set of objects in a pre-confluence
   - What is the structure of the set of closed elements ?
   - How is this structure related to FCA ?

Plan

Some terminology (lost in translation from data mining)

- Element of a partially ordered language L = pattern = motif
- attribute, property = item
- an object o = entry in a database
- $ext(t) = support(t)$ w.r.t. a database O

Support-closed motifs

Let F be a pattern language

Definition (Support-closed motifs)

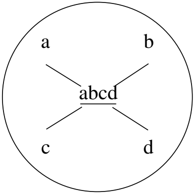
- $t \equiv_O t'$ iff $ext(t) = ext(t')$
- The maximal elements of the equivalence classes are the support-closed elements.

Are support-closed elements obtained as closed elements w.r.t some closure operator on F ?

Example

Objects are described as elements of $2^{\{a,b,c,d,e\}}$
$F \subseteq 2^{\{a,b,c,d,e\}}$, $O = \{o\}$
$d(o) = abcd$

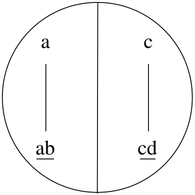
$F = 2^{\{a,b,c,d,e\}}$



$f(a) = abcd$

FCA

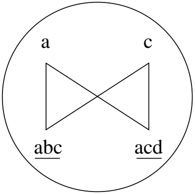
$F = \{a, c, ab, cd, abcde\}$



$f(a) = ab$
$f(c) = cd$

F is a preconfluence

$F = \{a, c, abc, abd, abcde\}$



$f(a) =?abc$
$=?abd$

No closure operator

Plan

Closure subsets of a finite lattice

Definition

Let E be an ordered set and $f : E \to E$ such that for any $x, y \in E$, $x \leq y \implies f(x) \leq f(y)$, $f(f(x)) = f(x)$, then:

- If $f(x) \geq x$, f is a closure operator
- If $f(x) \leq x$, f is a dual closure operator or a projection

If f is a closure operator, any $x = f(x)$ is called a closed element

A closure subset is the range $f[E]$ of some closure operator f.

Theorem (T.S. Blyth)

Let T be a finite lattice.

- *$C \subseteq T$ is a closure subset if and only if C is closed under meet.*
- *$A \subseteq T$ is a dual closure subset if and only if A is closed under join.*

A dual closure subset of T is associated to a projection and is also called an abstraction.

Pre-confluences

Definition

Let F be an ordered set s.t.

- *for any $t \in F$, $\uparrow t$ is a $\wedge$-semilattice and has a top element.*

F is a pre-confluence, $x \wedge_t y$ is a local meet

Lemma

Let F be a pre-confluence, then for any t in F and $x, y \in \uparrow t$

1. *$\uparrow t$ is a lattice with as join $x \vee_F y$, the least element of $\uparrow x \cap \uparrow y$*

Lemma

F is a pre-confluence if and only if for any $m \in \min(F)$, $\uparrow m$ is a $\wedge$-semilattice and has a top element.

A lattice is a pre-confluence with an infimum

Closure subsets of a pre-confluence

Definition

A subset C of a pre-confluence F is said closed under local meet whenever for any t in F, c_1, c_2 in C s.t. $c_1 \geq t$ and $c_2 \geq t$, we have

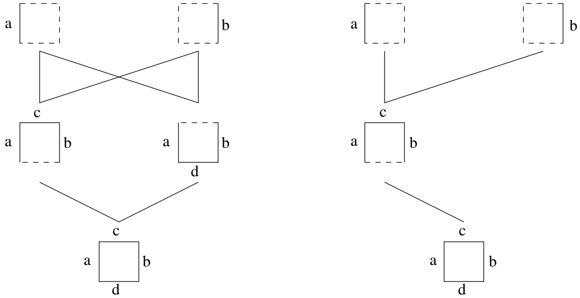
$$c_1 \wedge_t c_2 \quad \textit{belongs to } C.$$

Theorem

Let F be a pre-confluence.

- *$C \subseteq F$ is a closure subset iff C is closed under local meet.*
- *C is a pre-confluence.*

An example



A family F of connected subgraphs each generated by a subset of the edges $\{a, b, c, d\}$ of the original graph.

- $a = abc \wedge_a abd$, $b = abc \wedge_b abd$
- $C = \{a, b, abc, abcd\}$ is closed under local meet and f is Identity except that $f(abd) = abcd$, the least element of C greater than or equal to abd.

Plan

The lattice case

Proposition (Diday & Emilion, 2003)

Consider the following conditions:

- *T is a finite lattice*
- *For any $o \in O$, there is a greatest element $d(o)$ in T among those whose extension contains o*

then, the pair (int, ext) where

$$int(e) = \bigwedge_{\{o \in e\}} d(o)$$

is a Galois connection.

$int(e)$ is the intension of e, $f : f(t) = int \circ ext(t)$ is a closure operator.
The pairs $(f(t), ext(t))$ form a lattice isomorphic to $f[T]$
Same as the Pattern Structures framework

The pre-confluence case

Proposition

Consider the following conditions:

- F *is a pre-confluence*
- *For any* $o \in O$ *and any* t *in* F *s.t* $o \in ext(t)$*, there is a greatest element* $d_t(o)$ *in the up set* $\uparrow t$ *among those whose extension contains* o

then, for any t *in* F*, the pair* (int_t, ext) *where*

$$int_t(e) = \bigwedge_{t \; \{o \in e\}} d_t(o)$$

is a Galois connection.

$int_t(e)$ is a local intension of e, $f : f(t) = int_t \circ ext(t)$ is a closure operator.

The pairs $(f(t), ext(t))$ form a pre-confluence isomorphic to $f[F]$

Plan

Objects are described as elements of a lattice $T \supseteq F$

Let T be a lattice, and $F \subseteq T$. Typically $T = 2^P$.
An object is described as an element of T, and any element of T may represent an object.
If $d(o) = x$, $o \in ext(t)$ rewrites as $x \geq t$

Proposition

Let F be a subset of T, these conditions are equivalent:

1. *For any x in T and any t in F s.t $x \geq t$, there is a greatest element $p_t(x)$ in the up set $\uparrow t$ (in F) among those smaller than or equal to t.*
2. *F is a pre-confluence with join $\vee_F = \vee$*

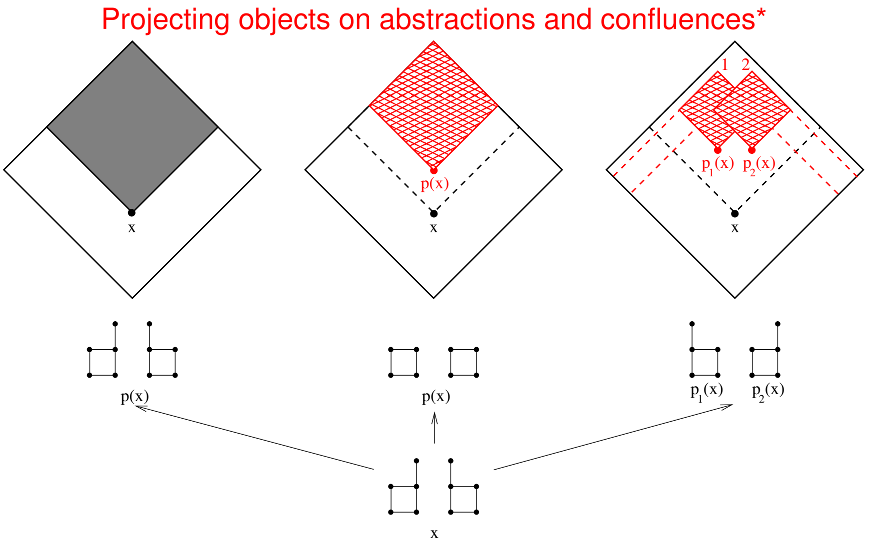
F is then denoted as a confluence of T*

Lemma

p_t is a projection on $T^t = \{y \in T | y \geq t\}$

An abstraction is a confluence* with an infimum

From lattices to confluences*



Galois pre-confluence as a union of concept lattices

This leads to a result close to the result of Boley and Coll.

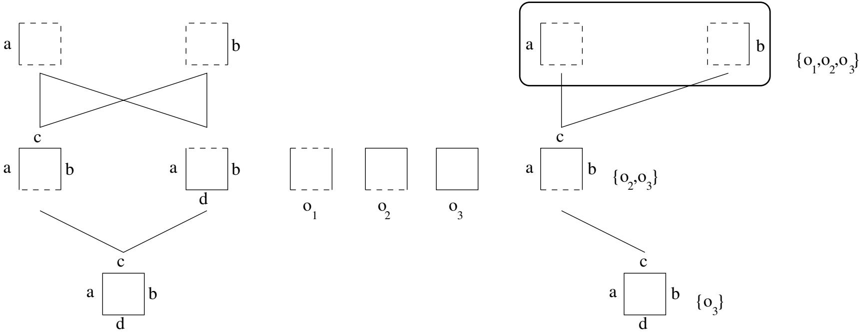
Proposition

Let $F \subseteq T$, the closure operator on F with respect to any set O of objects described in T exists if and only if F is a confluence.*

- The closure operator is defined by $f(t) = p_t \circ int \circ ext(t)$
- As p_t is a projection, $(p_t \circ int, ext)$ defines a projected concept lattice on $\uparrow t$ (Pernelle et al, 2002, Ganter and Kuznetsov, 2001).
- The closure subset $f[F]$ rewrites then as

$$\bigcup_{m \in min(F)} f[\uparrow m]$$

where each $f[\uparrow m]$ is the closure subset isomorphic to the projected concept lattice associated to p_m.



On the left the pre-confluence F
On the right the Galois pre-confluence $f[F]$ with respect to $\{o_1, o_2, o_3\}$.

- *abc* and *abcd* are the greatest connected subgraphs whose extensions are respectively $\{o_2, o_3\}$ and $\{o_3\}$.
- a and b are the bottom elements of $f[\uparrow a]$ and $f[\uparrow b]$.

The box around a and b means that a and b have the same extension $\{o_1, o_2, o_3\}$.

Plan

Extensional abstractions on confluences*

An extensional abstraction is a dual closure subset A of 2^O and leads to an extensional abstract (concept) lattice in which extensions are restricted to elements of A (Soldano & Ventos, 2011).
Extensional abstract pre-confluences are defined in the same way:

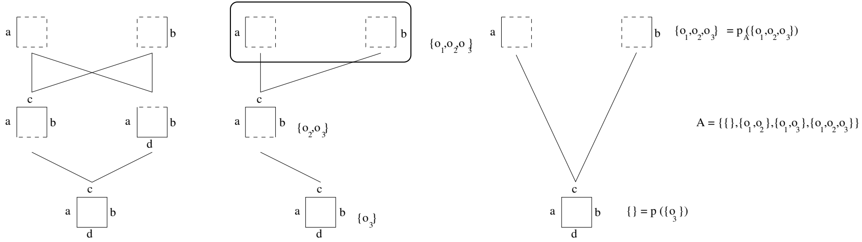
Proposition

Let F be a confluence of a lattice T, $A = p_A(2^O)$ an abstraction of 2^O, then:*
$f_A(t) = p_t \circ int \circ p_A \circ ext(t)$ is a support closure operator on F with respect to A and $f_A[F]$ is a pre-confluence.

$\Rightarrow$
Abstract closed patterns exists in confluences*

Example

F is represented on the left, $O = \{o_1, o_2, o_3\}$,
$A = \{\emptyset, \{o_1, o_2\}, \{o_1, o_3\}\{o_1, o_2, o_3\}\}$
with $d(o_1) = ab, d(o_2) = abc, d(o_3) = abcd$
The abstract Galois pre-confluence $f_A[F]$ is displayed on the right



$p_A \circ ext(a) = p_A \circ ext(b) = \{o_1, o_2, o_3\}$ as in the non abstract case, but
$p_A \circ ext(abc) = p_A(\{o_2, o_3\}) = \emptyset$ and therefore $f_A(abc) = abcd$

Plan

Implication bases, algorithmics . . .

$p \to q$ holds on O iff $ext(p) \subseteq ext(q)$

Definition (Min-max basis of implications)

Let F be a confluence and f the closure operator w.r.t. O*

- *The min-max basis $B = B_i \cup B_e$ of implications in F is the set $\{p \to q \mid ext(p) = ext(q) = e, p \neq q, p \text{ is minimal}, q \text{ is closed}\}$*
- *The internal sub basis B_i contains $p \to q$ where $p \leq q$*
- *The external sub basis B_e contains $p \to q$ where $\{p, q\}$ are unordered.*

Example

- Internal implication $1 - 2 - 3 \to 1 - 2 - 3 - 4$
- External implication $1 - 2 - 3 \to 7 - 8 - 9$

PARAMINER (Negrevergne et al, 2014) computes efficiently frequent closed patterns in *strongly accessible* confluent families (between two elements of F there must be a path of attributes within F).

Conclusion

We have reported that support-closure operators on partial orders F relying on local meet operators exist and we have connected to FCA by

- extending the results on closure subsets to pre-confluences
- showing that support-closed elements form a pre-confluence made of a union of concept lattices
- showing that extensional abstraction applies to Galois pre-confluences

- Different from defining closure operators on 2^F (as in (Kuznetsov and Samokhin, 2005) or on mappings (see recent work (J. Medina-Moreno and Coll. 2013) on multilattices).

- Many open computational and formal problems (building diagrams, implication basis construction, . . .)

- and more philosophical question about concepts (one extent, several intents ?)