Horn Minimization

An overview of some existing algorithms

Simon Vilmin

HSE - ISIMA

June 5, 2018

Introduction

- correlation relations (*implications*)

  e.g: movies and genres, *cyber-punk* $\longrightarrow$ *sci-fi*

- minimization without loss of knowledge

  e.g: *sci-fi* $\longrightarrow$ *sci-fi* is useless

- *study* of existing algorithms

Outline

I - Horn theories

Closure and implications

Minimization task

II - Some Algorithms

Minimizing the input

Building the result

III - Experiments

Experiments

Closure operator and systems
1.1 - Closure and implications

NATIONAL RESEARCH UNIVERSITY

Set Σ of attributes. A map $\varphi : 2^{\Sigma} \longrightarrow 2^{\Sigma}$ is a *closure operator* if, $\forall X, Y \subseteq \Sigma$:

- $X \subseteq \varphi(X)$ *(increasing)*
- $X \subseteq Y \longrightarrow \varphi(X) \subseteq \varphi(Y)$ *(isotone)*
- $\varphi(\varphi(X)) = \varphi(X)$ *(idempotent)*

Associated terminology:

- X is *closed* if $X = \varphi(X)$,
- Σ^{φ} set of closed sets: *closure system*.
- Σ^{φ} is closed under *intersection*, contains Σ.

NATIONAL RESEARCH UNIVERSITY

$A, B \subseteq \Sigma$. An *implication* is:

- $A \longrightarrow B$, A *premise*, B *conclusion*.
- $M \subseteq \Sigma$ *model* of $A \longrightarrow B$
  - $B \subseteq M \vee A \nsubseteq M \quad (\simeq B \vee \neg A)$
  - $M \models A \longrightarrow B$, $A \longrightarrow B$ follows from M.

Set of implications $\mathcal{L}$: *implication system*.

- $\mathcal{L} \models A \longrightarrow B$: all models of $\mathcal{L}$ are models of $A \longrightarrow B$,
- $\mathcal{L} \models A \longrightarrow B$ iff $B \subseteq \mathcal{L}(A)$.

Implications and closure

1.1 - Closure and implications

$\mathcal{L}$ an implication system:

- models of $\mathcal{L}$ form a *closure system*:

$$\Sigma^{\mathcal{L}} = \{M \subseteq \Sigma \mid M = \mathcal{L}(M)\}$$

- *closure operator* $\mathcal{L}(X)$: *smallest model* (inclusion wise) of $\mathcal{L}$ containing X, $X \subseteq \Sigma$.

$$\mathcal{L}(X) = \bigcap\{M \in \Sigma^{\mathcal{L}} \mid X \subseteq M\}$$

Small Example

1.1 - Closure and implications

Let $\mathcal{L}$ be an implication system:

- $\Sigma = \{a,\ b,\ c,\ d,\ e,\ f\}$,
- $\mathcal{L} = \{ab \longrightarrow cde, cd \longrightarrow f, c \longrightarrow a, d \longrightarrow b, abcd \longrightarrow ef\}$.

We have:

- $\mathcal{L}(b) = b$, b is *closed*, hence a *model* of $\mathcal{L}$,
- $\mathcal{L}(ab) = abcdef$, ab is *not* a model $\quad$ ($abcdef$ *is*),
- $\emptyset$ is also a model.

Redundancy, equivalence

1.2 - Minimization task

NATIONAL RESEARCH UNIVERSITY

$\mathcal{L}$ and $\mathcal{L}'$ implication systems:

- $A \longrightarrow B \in \mathcal{L}$ *redundant* if $\mathcal{L} - \{A \longrightarrow B\} \models A \longrightarrow B$,
- $\mathcal{L}' \models \mathcal{L}$: all implications of $\mathcal{L}$ follow from $\mathcal{L}'$,
- $\mathcal{L}' \models \mathcal{L}$ and $\mathcal{L} \models \mathcal{L}'$: *equivalent* bases.

Remind property: $(\mathcal{L}^- := \mathcal{L} - \{A \longrightarrow B\})$

- $\mathcal{L}^- \models A \longrightarrow B$ iff $B \subseteq \mathcal{L}^-(A)$

Minimum basis

1.2 - Minimization task

$\mathcal{L}$ *minimum* if no possible $\mathcal{L}'$ such that:

- $\mathcal{L}'$ equivalent to $\mathcal{L}$,
- $\mathcal{L}'$ has fewer implications than $\mathcal{L}$.

Particular minimum basis:

- $P \subseteq \Sigma$ *pseudo-closed* in $\mathcal{L}$ if:
  - $P \neq \mathcal{L}(P)$,
  - if $Q \subset P$ and Q pseudo-closed, then $\mathcal{L}(Q) \subseteq P$.
- *Duquenne-Guigues* (*canonical*) base:

$$\{P \longrightarrow \mathcal{L}(P) \mid P \textit{ pseudo-closed} \}$$

Small Example

1.2 - Minimization task

NATIONAL RESEARCH UNIVERSITY

Let $\mathcal{L}$ be an implication system:

- $\Sigma = \{a,\ b,\ c,\ d,\ e,\ f\}$,
- $\mathcal{L} = \{ab \longrightarrow cde, cd \longrightarrow f, c \longrightarrow a, d \longrightarrow b, abcd \longrightarrow ef\}$.

We have:

- $abcd \longrightarrow ef$ is *redundant*
- $cd \longrightarrow f$ can be deleted, if $ab \longrightarrow cde$ replaced by $ab \longrightarrow cdef$
- canonical basis: $ab \longrightarrow abcdef, c \longrightarrow ac, d \longrightarrow bd$

Tools

2.1 - Minimizing the input

Principle:

- reducing an input base $\mathcal{L}$,
- algorithms from FCA and databases,
-

First algorithm

2.1 - Minimizing the input

NATIONAL RESEARCH UNIVERSITY

MINCOVER:

- from Day, Wild, Duquenne, Guigues (80's)
- two steps:
  1. *maximize* the knowledge
     $A \longrightarrow B$ becomes $A \longrightarrow \mathcal{L}(A)$
  2. *remove* redundant information
     if $\mathcal{L}(A) = \mathcal{L}^{-}(A)$, $A \longrightarrow \mathcal{L}(A)$ is *redundant*
- output: *canonical base*.

Recall: $\mathcal{L}^{-} = \mathcal{L} - \{A \longrightarrow \mathcal{L}(A)\}$

A variation

2.1 - Minimizing the input

DuquenneMin:

- variation of MinCover by Duquenne, 2007,
- two steps:
  1. *left-saturation* and *redundancy* elimination
     if $B \subseteq \mathcal{L}^{-}(A)$, $A \longrightarrow B$ is useless
     else $A \longrightarrow B$ becomes $\mathcal{L}^{-}(A) \longrightarrow (\mathcal{L}^{-}(A) \cup B)$
  2. *iteratively build* pseudo-closed implications base
     if $\mathcal{L}^{-}(A)$ *pseudo-closed*, add $\mathcal{L}^{-}(A) \longrightarrow \mathcal{L}(A)$ to the result
- output: *canonical base*.

Database approach

2.1 - Minimizing the input

NATIONAL RESEARCH UNIVERSITY

MAIERMIN:

- functional dependency based algorithm, Maier, 80's
- steps:
  1. *redundancy* elimination
     if $B \subseteq \mathcal{L}^{-}(A)$, remove $A \longrightarrow B$
  2. *equivalence classes* reduction
     group implications by premises *closure*
     reduce those classes
- output: minimum basis

General principle

2.2 - Building the result

Query learning and Angluin algorithm:

- aim: learn a theory by constructing an *hypothesis*
- *oracle* answering *queries* (questions)
- queries:
  - *equivalence*: is our hypothesis equivalent to the target?
  - *membership*: is a set a model of the target?
- improve by *counterexample*
  - *positive*: model of the target, not of the hypothesis
  - *negative*: model of the hypothesis, not of the target

In our case

2.2 - Building the result

- no oracle to avoid *randomness*,
- no positive counter-example,
- premises are sufficient to have negative counter-example.

AFP Algorithm

2.2 - Building the result

NATIONAL RESEARCH UNIVERSITY

BERCZIMIN:

- logic based, Berczi, 2017
- principle:
  1. *build* a basis $\mathcal{L}_c$ against the input $\mathcal{L}$
  2. repeat minimality selection up to *equivalence*
  3. *minimality selection*:

     select the next minimal negative counter-example A,
     add $A \longrightarrow \mathcal{L}(A)$ to $\mathcal{L}_c$
- output: *canonical* base.

NATIONAL RESEARCH UNIVERSITY

Practical aspect:

- context of FCA, previous study of closure operators,
- use datasets from UCI repository,
- C++, boost.

Theoretical considerations

- using CLOSURE,
-

Overhaul results
3.1 - Experiments

$\mathcal{L}$		MinCov	Duq	Maier	Berczi	AFP
Flare	DG	0.097	0.117	0.211	27.922	115.656
	min 1	0.134	0.194	0.288	27.750	116.594
	min 2	0.200	0.190	0.308	30.063	108.812
	proper	1.684	0.933	0.917	88.375	524.031
	mingen	16.047	7.981	7.576	160.328	2810.620
Breast Cancer	DG	0.089	0.109	0.203	33.047	90.031
	min 1	0.108	0.126	0.243	26.578	89.516
	min 2	0.153	0.132	0.258	29.438	105.141
	proper	1.920	0.864	1.014	93.266	429.844
	mingen	2.006	1.277	1.444	102.562	598.172

Table: Comparison of the algorithms on real datasets (execution in s)

Observation

3.1 - Experiments

Observations *on these data*:

- cost of AFP and BERCZIMIN,
- DUQUENNEMIN, MAIERMIN efficient and similar on *non-minimum cases*,
- MINCOVER better on right-closed minimum cases.

Test observations:

- random generation
-

Some more results

3.1 - Experiments

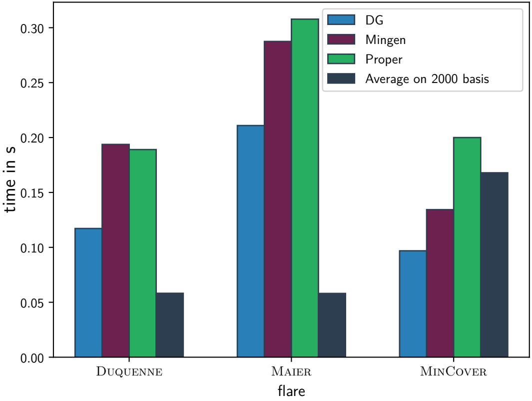

Figure: Flare - Random against and minimum basis: 49 attr, 3382 imp

Some more results

3.1 - Experiments

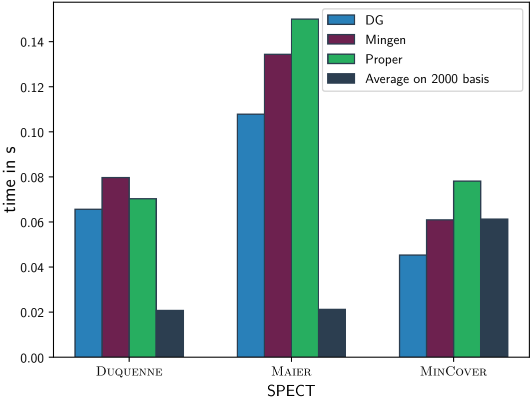

Figure: SPECT - Random against minimum basis: 23 attr, 2169 imp

NATIONAL RESEARCH UNIVERSITY

Observations:

- results of MinCover: proper basis is worse,
- redundant cases: Duquenne, Maier more efficient.

Explanations:

- *redundancy elimination* as first step,
- *right-closeness* of basis,
- suggests a study of underlying *structure*.

Boundaries:

- valid in *our context*,
- random generation,
- suggests *extension* of tests.

Conclusion

Purpose:

- *study* of minimization algorithms.

Results:

- algorithms from various communities,
- in practice: *redundancy elimination*

Perspectives:

- *theoretical* study of systems structure, AFP proof and complexity,
- *experimental* aspect, extend tests.