Readme: Horn Minimization

This repository contains implementation of Horn minimization algorithms. It relies on the code of https://github.com/yazevnul/fcai for FCA and closure testing purposes. In the first section we provide a brief theoretical recall of the minimization task: its terminology (in closure systems and implications), its algorithms and so forth. In the second part however, we will be interested in implementation details, uses, possible improvements and so on.

Appetizers: a bit of theory

Some definitions

The reader is expected to have background in set theory, and *Formal Concept Analysis* related mathematical fields (see **[9, 10]**). Let Σ be a set (of attributes).

> **Definition 1:** An application $\varphi : \Sigma \mapsto \Sigma$ is a closure operator iff for all $X, Y \subseteq \Sigma$:
>
> 1. $X \subseteq \varphi(X)$ (*extensive*),
> 2. $X \subseteq Y \implies \varphi(X) \subseteq \varphi(Y)$ (*monotone*),
> 3. $\varphi(X) = \varphi(\varphi(X))$ (*idempotent*).

To each closure operator φ over Σ is associated a closure system Σ^{φ} as in the next definition.

> **Definition 2:** A family of set $\Sigma^{\varphi} \subseteq 2^{\Sigma}$ is a closure system if and only if:
>
> 1. $\Sigma \in \Sigma^{\varphi}$,
> 2. $\bigcap_{s \in \mathcal{S}} s \in \Sigma^{\varphi}, \quad \forall \mathcal{S} \subseteq \Sigma^{\varphi}$.

Now let us focus on implications, being our main object.

> **Definition 3:** An implication $A \longrightarrow B$ over Σ is a pair of subsets $A, B \subset \Sigma$.

> **Definition 4:** A subset M of Σ is a model for some implication $A \longrightarrow B$ if $A \nsubseteq M \vee B \subseteq M$. In such case, we write $M \models A \longrightarrow B$.

A set $\mathcal{L}$ of implications is called an implication base . It is similar to *Horn formulae* of propositional logic, hence the Horn minimization name. Of course, every implication of $\mathcal{L}$ is defined over the same Σ. A set $M \subseteq \Sigma$ is a model for $\mathcal{L}$ if it is a model for all implications of $\mathcal{L}$. The set of models of $\mathcal{I}$ is a closure system, $\Sigma^{\mathcal{L}}$, and the corresponding closure operator $\mathcal{L}(\dot{})$ associate to $X \subseteq \Sigma$ the smallest model of $\Sigma^{\mathcal{L}}$:

$$\mathcal{L}(X) := \bigcap\{A | X \subseteq A, A \in \Sigma^{\mathcal{L}}\}$$

Two bases are equivalent if they have the same set of models. This leads to a definition of minimality.

> **Definition 5:** $\mathcal{L}$ is minimum if there is no equivalent base $\mathcal{L}'$ with fewer implications.

In particular, the Duquenne-Guigues (or canonical) base, defined as (see **[11]**):

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

is minimum. To be precise, let us finally remind what is pseudo-closedness.

> **Definition 6:** A set $P \subseteq \Sigma$ is a pseudo-closed set with respect to $\mathcal{L}(\dot{})$ if:
>
> 1. $P \neq \mathcal{L}(P)$,
> 2. for $Q \subset P$, if Q is pseudo-closed then $\mathcal{L}(Q) \subseteq P$.

To conclude this section, let us observe some examples. First, let $\Sigma = \{a, b, c\}$ and \$\mathcal{L} = { ab \longrightarrow c, c \longrightarrow b }. The corresponding set of models is $\{\emptyset, b, a, bc, abc\}$. This closure system can be represented as a Hasse diagram, notably because it is a complete lattice equipped within $\subseteq$ as order:

```
            abc
             o
            / \          Models of L, ordered
           /   \             by set-inclusion.
          /     \
     bc  o       \
         |        o a
     b   o       /
          \     /
           \   /
            \ /
             o
             ∅
```

Another example to consider, $\Sigma = \{a, b, c, d, e, f\}$ and $\mathcal{L} = \{ab \longrightarrow cde, cd \longrightarrow f, c \longrightarrow a, d \longrightarrow b, abcd \longrightarrow ef\}$. This basis is not minimum, some implications are redundant : they can be removed without altering the knowledge they depict. This is the case for the last one. The canonical basis here will be

$$\mathcal{L}_c = \{ab \longrightarrow abcdef, c \longrightarrow ac, d \longrightarrow bd\}$$

Algorithms

We implemented 5 algorithms:

- `MinCover` : an algorithm found in **[7, 9, 14]**,
- `DuquenneMin` : being a variation of `MinCover` found in **[8]**,
- `MaierMin` : a procedure coming from database community, see **[12, 13]**,
- `BercziMin` : based on hypergraph representation of logical Horn formula (**[6]**)
- `AFP` : an algorithm derived from query learning, further details in **[1, 2]**

For closure algorithms, please refer to **[5, 9]**

MinCover

Main dish: implementation

References

- **[1]** Angluin, D., Frazier, M., and Pitt, L. *Learning conjunctions of Horn clauses*. Machine Learning 9, 2 (July 1992), 147–164.
- **[2]** Arias, M., and Balcázar, J. L. *Canonical Horn Representations and Query Learning*. In Algorithmic Learning Theory (Berlin, Heidelberg, 2009), R. Gavaldà, G. Lugosi, T. Zeugmann, and S. Zilles, Eds., Springer Berlin Heidelberg, pp. 156–170.
- **[3]** Ausiello, G., D'Atri, A., and Saccà, D. *Graph Algorithms for Functional Dependency Manipulation*. J. ACM 30, 4 (Oct. 1983), 752–766.
- **[4]** Ausiello, G., D'Atri, A., and Saccá, D. *Minimal Representation of Directed Hypergraphs*. SIAM J. Comput. 15, 2 (May 1986), 418–431.
- **[5]** Beeri, C., and Bernstein, P. A. *Computational Problems Related to the Design of Normal Form Relational Schemas*. ACM Trans. Database Syst. 4, 1 (Mar. 1979), 30–59.
- **[6]** Bérczi, K., and Bérczi-Kovács, E. R. *Directed hypergraphs and Horn minimization*. Information Processing Letters 128 (2017), 32 – 37.
- **[7]** Day, A. *The Lattice Theory of Functional Dependencies and Normal Decompositions*. International Journal of Algebra and Computation (1992).
- **[8]** Duquenne, V. *Some variations on Alan Day's algorithm for calculating canonical basis of implications*. In Concept Lattices and their Applications (CLA) (Montpellier, France, 2007), pp. 17–25.
- **[9]** Ganter, B., and Obiedkov, S. *Conceptual Exploration*. Springer, 2016.
- **[10]** Ganter, B., and Wille, R. *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag, Berlin Heidelberg, 1999.
- **[11]** Guigues, J., and Duquenne, V. *Familles minimales d'implications informatives résultant d'un tableau de données binaires*. Mathématiques et Sciences Humaines 95 (1986), 5–18.
- **[12]** Maier, D. *Minimum Covers in Relational Database Model*. J. ACM 27, 4 (1980), 664 – 674.
- **[13]** Maier, D. *Theory of Relational Databases*. Computer Science Pr, 1983.
- **[14]** Wild, M. *Implicational bases for finite closure systems*. Informatik-Bericht 89/3, Institut fuer Informatik (Jan. 1989).
- **[15]** Wild, M. *A Theory of Finite Closure Spaces Based on Implications*. Advances in Mathematics 108, 1 (Sept. 1994), 118–139.
- **[16]** Wild, M. *Computations with finite closure systems and implications*. In Computing and Combinatorics (Aug. 1995), Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 111–120.