Utilisation des irréductibles d'un treillis de concepts pour la sélection de motifs

Michel Liquiere

Utilisation des irréductibles d'un treillis de concepts pour la sélection de motifs

Michel Liquière

LIRMM, Université Montpellier 2, CNRS, Montpellier, France

Résumé. Nous présentons une méthode de sélection de motifs basée sur l'analyse des concepts formels de Wille (1982). Elle repose sur la notion $\wedge$-irréductibles d'un treillis de concepts. Un algorithme incrémental, sur l'ajout ou le retrait d'attributs, qui maintient la liste des $\wedge$-irreductibles est ensuite proposé.

1 Introduction

Notre problématique générale est celle d'un système fournissant des motifs, à partir d'exemples ayant des descriptions complexes, qu'il faut sélectionner avec une approche formelle permettant une explication de la sélection.

Une des démarches possible, pour le traitement des données structurées mais aussi pour l'acquisition d'un langage de motifs pertinents, passe par une transformation de la description des exemples de structurelle à propositionnelle. Pour cela on recherche un ensemble de motifs pertinents et on les utilise comme un ensemble d'attributs booléens qui permettent donc une description propositionnelle des exemples. Le choix de ces attributs est bien sûr crucial.
Pour avoir une approche formelle à ce problème nous allons utiliser l'analyse des concepts formels (Formal Concept Analysis et en abrégé *FCA*) qui a été formalisée par Wille (1982). L'analyse des concepts formels a, en effet, le mérite d'offrir un cadre théorique et une définition précise de la notion de concepts, ces concepts s'organisant dans une structure de Treillis. Il est connu que dans un treillis il existe des éléments particuliers nommés irréductibles (Medina-Moreno et al., 1308), (Li et al., 2013). Comme dans le cas de l'analyse des concepts formels l'espace de recherche est un treillis, cette notion d'éléments irréductibles existe aussi et permet d'identifier les motifs essentiels (au sens de nécessaires) aux regroupements/séparations des exemples.

2 Analyse des concepts formels

Dans cette section nous donnons différentes définitions pour arriver aux définitions d'éléments irréductibles dans le cadre de l'analyse des concepts formels.

Definition 1 Un ensemble ordonné $\mathscr{L}$ est un treillis $\Leftrightarrow$ pour deux éléments x et y le supremum $x \vee y$ et l'infinum $x \wedge y$ existent.

Un treillis est complet si pour tout sous ensemble X et Y de $\mathscr{L}$, $\bigvee X$ et $\bigwedge Y$ existent dans $\mathscr{L}$.

Definition 2 $\wedge$-irréductible
Soit (T,$\leq$) un treillis fini, un élément t de T est dit $\wedge$-irréductible $\Longleftrightarrow$:
- $t \neq \top$ (l'élément minimal du treillis) et
- t= a $\wedge$ b => t=a ou t=b

Proposition 1 *Un élément a d'un treillis fini est $\wedge$-irréductible $\Leftrightarrow$ il a exactement un seul voisin supérieur.*

Cette proposition exprime le fait qu'un élément t est $\wedge$-irréductible si et seulement si t ne peut pas être représenté comme $\wedge$ (l'infinum) de deux éléments du treillis.

Proposition 2 *Dans un treillis fini, chaque élément est le $\wedge$ (resp. $\vee$) d'éléments $\wedge$-irréductibles (resp. $\vee$-irréductibles). Barbut et Monjardet (1970)*

L'ensemble des $\vee$-irréductibles (reps. $\wedge$-irréductibles) est important car, à partir de ces éléments on peut construire l'ensemble du sup-demi-treillis (resp.inf-demi-treillis) en utilisant l'opérateur $\vee$ (resp. $\wedge$) (à l'exception de l'élément maximal $\perp$ (resp. minimal $\top$)).

Nous donnons maintenant la définition de contexte formel provenant de l'analyse des concepts formels.

Definition 3 Contexte formel
On nommera contexte formel un triplet ($\mathscr{E}$,$\mathscr{A}$,I) où
- $\mathscr{E}$ est un ensemble fini nommé ensemble des objets (les exemples)
- $\mathscr{A}$ est un ensemble fini nommé ensemble des attributs (les propriétés)
- I une relation binaire liant les objets et les attributs.

Definition 4 Concept en FCA
Pour un contexte formel ($\mathscr{E}$,$\mathscr{A}$,I) un concept est un couple (E,A) avec
E $\subseteq \mathscr{E}$, A $\subseteq \mathscr{A}$ | E(A)=E et A(E)=A tel que

$A : 2^{\mathscr{E}} \rightarrow 2^{\mathscr{A}}$ et A(E)={ a $\in \mathscr{A}$ / $\forall$e$\in$E : e I a } (lire A(E) comme les attributs communs à l'ensemble des objets de E)

$E : 2^{\mathscr{A}} \rightarrow 2^{\mathscr{E}}$ et E(A)={ e $\in \mathscr{E}$ / $\forall$a$\in$A : eI a } (lire E(A) comme les exemples communs à l'ensemble des attributs de A)
Où l'on note :
- $2^{\mathscr{E}}$ l'ensemble des parties de l'ensemble $\mathscr{E}$.
- pour a $\in \mathscr{A}$ et e a $\in \mathscr{E}$, e I a $\Leftrightarrow$ (e,a) $\in$ I
- et pour A $\subseteq \mathscr{A}$ on notera F(A)= A(E(A)) fermeture de A (noté aussi A" par Wille)

A partir de cette définition nous avons le théorème fondamental de la FCA :

Theorem 1 *L'ensemble des concepts d'un contexte formel forme un treillis complet nommé treillis de concepts.*

Preuve [Ganter et Wille (1999)]

3 $\wedge$-irréductibles

Pour un contexte formel ($\mathscr{E}$,$\mathscr{A}$,I) l'ensemble $\mathscr{A}$ contient l'ensemble des éléments $\wedge$-irréductibles et chaque $\wedge$-irréductible est associé à un attribut (Ganter et Wille, 1999). On sait que pour un contexte formel si l'on retire l'ensemble de ses attributs non irréductibles on obtient, pour ce nouveau contexte, un treillis des concepts isomorphe au treillis du contexte initial. C'est à dire permettant les mêmes regroupements des exemples.
Cette propriété est aussi vraie si on n'enlève qu'un sous-ensemble des éléments réductibles (non irréductibles) et en particulier un sous ensemble des attributs réductibles.
Sachant que la construction du treillis de concepts donne l'ensemble des regroupements des exemples sans redondance, par rapport à la définition de concept, on voit que les attributs $\wedge$-irréductibles jouent le rôle de la base minimale de descriptions sans perte de possibilités quand aux regroupements des exemples (Ganter et Wille, 1999).

Suivant (Ganter et Wille, 1999) nous donnons maintenant une propriété fondamentale permettant de caractériser un attribut $\wedge$-irréductible dans le cadre d'un contexte formel.

Definition 5 Pour un contexte formel ($\mathscr{E}$,$\mathscr{A}$,I)

Pour un objet $e \in \mathscr{E}$ et un attribut $a \in \mathscr{A}$ on note $e \nearrow a \Leftrightarrow \neg\ (e\ I\ a)$ et pour tout attribut a_1 / $E(a) \subseteq E(a_1)$ et $E(a) \neq E(a_1) \Rightarrow (e\ I\ a_1)$

Proposition 3 *(Ganter et Wille, 1999)*

Pour un contexte (clarifié) ($\mathscr{E}$,$\mathscr{A}$,I), un attribut $a \in \mathscr{A}$ est $\wedge$-irréductible $\Leftrightarrow \exists\ e \in \mathscr{E}$ avec $e \nearrow a$
Un contexte est clarifié <=> $\forall\ a_1, a_2 \in \mathscr{A}$ on a $E(a_1) \neq E(a_2)$

On voit donc qu'un attribut est $\wedge$-irréductible s'il existe un ensemble non vide d'exemples qui ne vérifient pas cet attribut mais qui vérifient tous les attributs appartenant à sa fermeture.

Nous allons maintenant préciser ce point pour permettre l'énoncé d'un algorithme utilisant cette propriété.

Definition 6 Soit un contexte ($\mathscr{E}$,$\mathscr{A}$,I)
Pour un attribut $a \in \mathscr{A}$ on nomme irr(a)= $\{e_i$ / $e_i \in \mathscr{E}$ et $\forall\ a_j \in F(a)$ avec E(a_j)$\neq$ E(a), (e_i I a_j)$\}$

Il s'agit de l'ensemble des exemples permettant de prouver que a est un $\wedge$-irréductible.

Proposition 4 $e \nearrow a \Leftrightarrow$ *irr(a)* $\neq \emptyset$

PROOF Vient directement de la proposition 3

Cette propriété montre que tout attribut $\wedge$-irréductible a est caractérisé par son ensemble d'exemples irr(a). Les exemples de irr(a) ne vérifient pas l'attribut a mais vérifient tous les attributs b avec e(a) $\subset$ e(b). En fait l'attribut a est fondamental pour la séparation des exemples de irr(a) des autres exemples.

La proposition 4 donne l'algorithme Rechercheirréductible suivant qui permet de calculer, pour un attribut a, irr(a) et donc de savoir si a est $\wedge$-irréductible :

Cet algorithme reçoit comme paramètres un attribut a et un contexte et retourne l'ensemble irr(a). Si le résultat de cet algorithme est un ensemble non vide alors a est un $\wedge$-irréductible. Pour réduire sa complexité cet algorithme utilise (en 7 :) la proposition 5 pour vérifier si E(a_i) = E(a) quand $a_i \in F(a)$ (la fermeture de a)

Algorithm 1 Recherche de irr(a)

procedure RECHERCHEIRRÉDUCTIBLE(a,($\mathscr{E}$,$\mathscr{A}$,I)) ▷ attribut a et un contexte formel
  $F(a) \leftarrow$ fermeture(a,($\mathscr{E}$,$\mathscr{A}$,I)) ▷ recherche A(E(a))
  E_t= E - E(a)
  for all $a_i \in F(a)$ **do**
    if $E_t = \emptyset$ **then**
      return E_t
    if |E(a_i)| $\neq$ |E(a)| **then**
      $E_t \leftarrow E_t \bigcap E(a_i)$
    else ▷ cas égalité entre atttributs voir proposition 5
      fusion a_i ▷ Sauve l'information sur l'égalité
  return E_t

Proposition 5 *E(a_i) = E(a)* $\Leftrightarrow a_i \in F(a)$ *et* | *E(a_i)*| = |*E(a)*|

PROOF $\Rightarrow$ E(a_i) = E(a) $\Rightarrow a_i \in F(a)$ et | E(a_i)| = |E(a)| $\Leftarrow a_i \in F(a) \Rightarrow$ E(a_i) $\supseteq$ E(a) comme | E(a_i)| = |E(a)| $\Rightarrow$ E(a_i) = E(a)

Complexité : Sachant que l'intersection de listes triées de taille m,n peut être faite en O(m+n)) on obtient comme complexité, dans le pire des cas, $O(|A| \times |E| + |E|)$. L'opération de fermeture ayant la même complexité.
Donc pour un contexte ($\mathscr{E}$,$\mathscr{A}$,I) la complexité de la recherche de l'ensemble des $\wedge$-irréductibles est $O(|A|^2 \times |E|)$

4 Ajout incrémental d'attributs et calcul de l'ensemble des $\wedge$-irréductibles

Dans le cas où un contexte évolue par ajout/retrait d'attribut un calcul complet de l'ensemble des $\wedge$-irréductibles, à chaque modification de l'ensemble des attributs, est couteux (bien que polynomial). Dans cette section nous étudions quelles sont les modifications à apporter à l'ensemble des $\wedge$-irréductibles dans le cas de l'ajout ou du retrait d'un attribut.

4.1 Ajout d'un attribut

On considère que pour un contexte ($\mathscr{E}$,$\mathscr{A}$,I) on connait, à l'étape n, l'ensemble de ses $\wedge$-irréductibles noté Irre($\mathscr{E}$,$\mathscr{A}$,I) et que pour chaque attribut a on a mémorisé $irr(a)$ c.a.d l'ensemble des exemples qui caractérise le fait que cet attribut est irréductible.
De plus dans le contexte ($\mathscr{E}$,$\mathscr{A}$,I) tous les attributs égaux ont été fusionnés. c.a.d que l'on a un attribut qui représente un ensemble d'attributs égaux.

Proposition 6 *Ajout d'un attribut a^n à un contexte ($\mathscr{E}$,$\mathscr{A}$,I) dont l'ensemble Irre($\mathscr{E}$,$\mathscr{A}$,I) des $\wedge$-irréductibles est connu.*

i) Si a^n est $\wedge$-irréductible et non égal à un attribut de A alors Irre($\mathscr{E}$,$\mathscr{A}+a^n$,I)= Irre($\mathscr{E}$,$\mathscr{A}$,I)+a^n

Et

ii) pour $a \in$ *Irre(* $\mathscr{E},\mathscr{A}$ *,I) si* $a^n \in F(a)$ *alors on modifie* $irr(a) = irr(a) \cap e(a^n)$

Si irr(a)=∅ alors Irre($\mathscr{E}$ *,Irre(* $\mathscr{E},\mathscr{A} + a^n$ *,I),I)= Irre(* $\mathscr{E},\mathscr{A}$ *,I)-* a

iii) pour $a \notin Irre($ $\mathscr{E},\mathscr{A}$ *,I) on a Irre(* $\mathscr{E},\mathscr{A} + a^n$ *,I)=Irre(* $\mathscr{E},\mathscr{A}$ *,I)*

PROOF i) Comme a^n est $\wedge$-irréductible alors il est simplement ajouté à l'ensemble des éléments $\wedge$-irréductibles.

ii) Comme un attribut a est $\wedge$-irréductible ssi irr(a)$\neq \emptyset$ L'ensemble irr(a) n'est modifié que via les attributs $a_i \in F(a)$ Si $a^n \notin F(a)$ alors pas de modification de irr(a) donc pas de modification de sa propriété d'être $\wedge$-irréductible. Sinon on modifie irr(a) et s'il devient vide alors a devient non $\wedge$-irréductible.

iii) Comme un attribut a non $\wedge$-irréductible possède un ensemble irr(a) vide et comme l'arrivée d'un nouvel attribut ne pourrait que potentiellement réduire cet ensemble (qui est produit par intersection) alors l'attribut a reste non $\wedge$-irréductible.

Pour l'ajout d'un nouvel attribut il faut :

- Vérifier si cet attribut est irréductible ou non : $O(|E| \times lg(|E|))$

- Pour l'ensemble des attributs $\wedge$-irréductible Irre(E,A,I) on recherche si $a^n \in F(a)$ pour cela on recherche donc si $e(a^n) \supseteq e(a)$ dont la complexité est $O(|E|)$ pour chaque attribut $\wedge$-irréductible. Au pire des cas $O(|Irre(E,A,I)| \times |E|)$ avec $|Irre(E,A,I)| \ll |A|$

Donc la complexité globale pour l'ajout d'un attribut est de : $O(|A| \times |E|) + |Irre(E,A,I)| \times |E|)$ qui est bien moindre que $O(|A|^2 \times |E|)$

4.2 Retrait d'un attribut

Proposition 7 *Retrait d'un attribut* a^p *à un contexte (* $\mathscr{E},\mathscr{A}$ *,I) dont l'ensemble Irre(* $\mathscr{E},\mathscr{A}$ *,I) des* $\wedge$*-irréductibles est connu.*

i) Si a^p *est* $\wedge$*-irréductibles alors Irre(* $\mathscr{E},\mathscr{A} - a^p$ *,I)= Irre(* $\mathscr{E},\mathscr{A}$ *,I)-* a^p

Et

ii) pour $a \in$ *Irre(* $\mathscr{E},\mathscr{A}, I)$ *si* $a^p \in F(a)$ *alors* a *reste* $\wedge$*-irréductibles mais on modifie* $irr(a)$ *qui peut augmenter.*

iii) pour $a \notin$ *Irre(* $\mathscr{E},\mathscr{A}$ *,I) si* $a^p \in F(a)$ *alors on modifie* $irr(a)$ *Si* $irr(a) \neq \emptyset$ *alors Irre(* $\mathscr{E},\mathscr{A} - a^p$ *,I)= Irre(* $\mathscr{E},\mathscr{A}$ *,I)+* a

PROOF i) Si a^p est $\wedge$-irréductible alors il est dans un premier temps simplement enlevé de l'ensemble des éléments $\wedge$-irréductibles.

ii) L'ensemble irr(a) est modifié que via les attributs $a_i \in F(a)$ Si $a^p \in F(a)$ alors irr(a) peut être modifié. Comme a^p est enlevé cela peut augmenter irr(a) qui est produit par une suite d'intersections. Donc on modifie irr(a) pour tous les attributs $a^p \in F(a)$.

Complexité

i) O(1)

ii) iii) On modifie $irr(a)$ pour tout attribut a $a^p \in F(a)$. La complexité de ces opérations est $O(|C| \times |A| \times |E|)$ ou C est l'ensemble des attributs tel que $a^p \in F(a)$.

5 Conclusion

Dans cet article, en nous basant sur une propriété donnée par (Ganter et Wille, 1999), nous avons présenté un algorithme permettant de trouver, en temps polynomial, si un attribut était $\wedge$-irréductible. Puis nous avons étendu cette approche à l'ajout ou le retrait d'un attribut dans un contexte formel et avons montré qu'en conservant certaines d'informations il était possible d'avoir une réduction de la complexité de l'approche par rapport à un calcul complet de l'ensemble des attributs $\wedge$-irréductibles. Il aurait été intéressant de pouvoir faire de même avec l'ajout ou le retrait d'exemple. Toutefois, dans ce cas nous n'avons pas trouvé de réductions de complexité par rapport au calcul complet de l'ensemble des $\wedge$-irréductibles.

Références

Barbut, M. et B. Monjardet (1970). *Ordre et classification. Algébre et combinatoire*. Hachette.

Ganter, B. et R. Wille (1999). *Formal Concept Analysis : Mathematical Foundations*. Springer-Verlag.

Li, T.-j., M.-z. Li, et Y. Gao (2013). Attribute reduction of concept lattice based on irreducible elements. *International Journal of Wavelets, Multiresolution and Information Processing 11*.

Medina-Moreno, J., M. E. Cornejo, et E. Ramírez (2013/08). Irreducible elements in multi-adjoint concept lattices. In *Proceedings of the 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13)*, pp. 125–131. Atlantis Press.

Wille, R. (1982). Restructuring lattice theory : An approach based on hierarchies of concepts. In I. Rival (Ed.), *Ordered Sets*, pp. 445–470. Reidel, Dordrecht-Boston.

Summary

We present a method of pattern selection based on the formal concepts analysis of Wille (1982). It is based on the $\wedge$-irreducible notion of a lattice of concepts. An incremental algorithm, based on the addition or removal of attributes, which maintains the list of $\wedge$-irreducible elements is then proposed.