Agrégation de systèmes de fermetures: cas des hiérarchies, topologies et géométries convexes.

Florent Domenach*

*Computer Science Department, Université de Nicosie, 46 Makedonitissas Av., 1700 Nicosie, Chypre, domenach.f@unic.ac.cy

1 Introduction

Nous considérons le problème où étant donné un k-uple $\mathcal{C}^* = (\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_k)$ de classifications du même type (hiérarchies, géométries convexes), on cherche une classification consensus, representative de $\mathcal{C}^*$. Lorsque la fonction consensus C opérant sur un k-uple $\mathcal{F}^* = (\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_k)$ de systèmes de fermeture d'un type particulier satisfait deux axiomes liés aux conditions d'Adams (Adams , 1986), on peut ajuster $C(\mathcal{F}^*)$ à un système de fermeture du même type (Domenach et Leclerc , 2007).

2 Cryptomorphismes de systèmes de fermeture

Soit S un ensemble fini, et $\mathcal{P}(S)$ l'ensemble de tous les sous-ensembles de S. Une classification peut être considérée comme une famille $\mathcal{C}$ de sous-ensembles (classes) sur S. On définit une famille $\mathcal{F} \subseteq \mathcal{P}(S)$ comme un *système de fermeture* sur S si S appartient à la famille et si elle est $\cap$-stable. Un cryptomorphisme des systèmes de fermeture sont les *systèmes implicatifs complets*, relations binaires sur $\mathcal{P}(S)$ satisfaisant $B \subseteq A$ implique $A \to B$, $A \to B$ et $B \to C$ impliquent $A \to C$, et $A \to B$ et $C \to D$ impliquent $A \cup C \to B \cup D$.

Un système de fermeture $\mathcal{F}$ est *hiérarchique* si $F, F' \in \mathcal{F}$ impliquent $F \cap F' \in \{\emptyset, F, F'\}$ et si $\{s\} \in \mathcal{F}$ pour tout $s \in S$, ou possède les propriétés suivantes sur son système implicatif : (IE) :$\forall A \subseteq S$, $A \neq \emptyset$, $\emptyset \not\to A$, (IS) : $\forall s \in S, A \subseteq S$, $\{s\} \notin A$, $\{s\} \not\to A$ et (IT) : $\forall A, B, C \subseteq S$, si $A \cap B \neq \emptyset$ alors $A \cup B \to C$ implique $A \to C$ ou $B \to C$. De même, $\mathcal{F}$ est une *topologie* si $F, F' \in \mathcal{F}$ impliquent $F \cup F' \in \mathcal{F}$ et si $\emptyset \in \mathcal{F}$, i.e. satisfaisant (IE) et (ID) : $s \in S, A \subseteq S$ et $A \to s$ impliquent $\exists a \in A : a \to s$. Enfin, $\mathcal{F}$ est une *géométrie convexe* si $\emptyset \in \mathcal{F}$ et si, pour tout $F \in \mathcal{F}$, il y a un plus petit (pour l'ordre d'inclusion) sous-ensemble B de S tel que $\varphi(B) = F$, ou si elle satisfait (IC) : $\forall A, B, C \subseteq S : A \subseteq C$ et $B \subseteq C$, $A \to B$ et $B \to C$ impliquent $A \cap B \to C$.

3 Consensus fréquents de systèmes de fermeture

Lorsqu'on considère l'agrégation d'un profil $\mathcal{F}^* = (\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_k)$ de systèmes de fermetures en un système de fermeture consensus $\mathcal{F}$, avec tous les $\mathcal{F}_i$ d'un type particulier, une méthode naturelle consiste à considérer les ensembles fermés présents dans "suffisament" de familles. Il est naturel d'envisager le *consensus p-fréquent* de $\mathcal{F}^*$, noté $F_{(p)}(\mathcal{F}^*)$, comme la famille de tous les ensembles p-fréquents, i.e. apparaissant dans au moins p familles.

Considérons la fonction consensus $E_{(p)}$ qui, à un profil de systèmes de classes, va prendre en compte les ensembles (et non plus les classes) présents dans suffisament de systèmes du profil. Formellement, on associe $E_{(p)}$ à un profil de systèmes de classes $\mathcal{F}^* = (\mathcal{F}_1, ..., \mathcal{F}_k)$ par $E_{(p)}(\mathcal{F}^*) = \bigcup_{J \subseteq K, |J| \geq p} \bigcap_{i \in J} \{A \not\to_i B\}$. $E_{(p)}$ peut être vue comme l'ensemble de toutes les paires $(A, B) \in \mathcal{P}(S) \times \mathcal{P}(S)$ qui sont liées par implications dans au plus p familles. Malheureusement, généralement $E_{(p)}$ n'est pas en correspondance avec un système de fermeture, et il nous faut considérer les deux conditions suivantes (Leclerc , 2007) : (RI1) : Pour tous $A, B \subseteq S$, $A \to B$ implique $|\{i \in K : A \to_i B\}| \geq k - p$ et (RI2) : Pour tout $F, F' \in \mathcal{F}$, $F \subset F'$ implique $|\{i \in K : F \to_i F'\}| < k - p$.

Lorsqu'on considère le système implicatif I satisfaisant (RI1) et (RI2), il apparait que $E_{(p)}$ satisfait (IE), (IS) ou (ID) si et seulement si I satisfait la même propriété. Toutefois, pour (IT) et (IC), il est intéressant de noter que nous sommes obligés de considérer la relation de consensus unanime $E_{(k)}$. On obtient ainsi le théorème d'Adams (la famille de hiérarchies $\mathcal{H} = E_{(k)}(\mathcal{H}^*)$ est l'unique hiérarchie satisfaisant (RI1) et (RI2)) et la proposition suivante :

Proposition : Si on considère des topologies T_i (resp. géométries convexes G_i) sur S, alors $E_{(p)}(\mathcal{T}^*)$ (resp. $E_{(k)}(\mathcal{G}^*)$) est l'unique topologie (resp. géométrie convexe) satisfaisant les conditions (RI1) et (RI2).

Références

Adams III, E.N. (1986). N-trees as nestings : complexity, similarity and consensus. *Journal of Classification 3*, 299-317.

Domenach, F. et B. Leclerc (2007). The structure of the overhanging relations associated with some types of closure systems. *Ann. Math. Art. Int. 49*, 137-149.

Leclerc, B. (2007). Consensus of classifications based on frequent groupings, in *Selected Contributions in Data Analysis and Classification*. Springer, 317-324.

Summary

We consider here the problem of adjusting systems of classes to a specific system, while focusing on specific cases (hierarchies, topologies and convex families). We will characterize a consensus function, defined on the complete implicational system associated with the systems of classes, by considering the presence of pairs of subsets linked by implication that can be found in at least p families which allows us to obtain a specific system.