Mathématiques et sciences humaines

Mathematics and social sciences

190 | Été 2010

Mathématiques discrètes : théories et usages.
Numéro en hommage à Bruno Leclerc

Consensus de familles de Moore typées

Consensus of particular Moore families

Florent Domenach

Édition électronique

URL : <http://journals.openedition.org/msh/11773>

DOI : 10.4000/msh.11773

ISSN : 1950-6821

Éditeur

Centre d'analyse et de mathématique sociales de l'EHESS

Édition imprimée

Date de publication : 10 mars 2010

Pagination : 125-137

ISSN : 0987-6936

Référence électronique

Florent Domenach, « Consensus de familles de Moore typées », *Mathématiques et sciences humaines* [En ligne], 190 | Été 2010, mis en ligne le 16 octobre 2010, consulté le 23 juillet 2020. URL : <http://journals.openedition.org/msh/11773> ; DOI : <https://doi.org/10.4000/msh.11773>

CONSENSUS DE FAMILLES DE MOORE TYPÉES

Florent DOMENACH¹

RÉSUMÉ – Soit $\mathcal{C}$ un ensemble de classes d'éléments d'un ensemble S . On considère classiquement qu'il va contenir S et qu'il va être stable par intersection, i.e. que c'est une famille de Moore. Cet article porte sur les possibilités d'ajustement d'un système de classes à une relation d'emboîtement (et donc d'implication) donnée. Pour tout entier p entre 1 et k , et pour tout profil de familles de Moore, on associe la fonction consensus créée par les paires de sous-ensembles de S emboîtées dans au moins p familles du profil. Nous allons montrer que cette fonction consensus permet d'obtenir une famille de Moore spécifique, et ce pour des profils de familles de Moore emboîtées, hiérarchiques, distributives, topologiques, pour des arbres de parties ou encore pour des géométries convexes.

MOTS CLÉS – Classification, Fermeture, Hiérarchie, Implication, Relation d'Emboîtement.

SUMMARY – Consensus of particular Moore families

Let $\mathcal{C}$ be a set of classes of elements on a set S . We usually consider that it will contain S and that it will be stable by intersection, i.e. it's a closure system. This article will look at the possibilities of fitting a system of classes to a specific "overhanging" relation (and so, to a specific implicational system). For any natural number p between 1 and k , we associate to any profile of closure systems the consensus mapping created by the pairs of subsets of S overhanged in at least p families of the profile. It will be shown that this consensus mapping allows us to create a specific closure system when we consider closure systems that are nested, hierarchical, distributive, trees of subsets, topologies or even convex geometries.

KEYWORDS – Classification, Closure Mapping, Hierarchy, Implication, Overhanging Relation.

1. INTRODUCTION

Le but de cet article est d'obtenir, en paraphrasant [Leclerc, Cucumel, 1987 ; Crown, Janowitz, 2003], une fonction consensus C opérant sur un k -uplet $(\mathcal{F}_1, \dots, \mathcal{F}_k)$ de « structures » d'un certain type et d'obtenir une « structure » $\mathcal{F}^c = C(\mathcal{F}_1, \dots, \mathcal{F}_k)$ qui d'une certaine manière résume ce que l'original k -uplet a en commun, tout en conservant l'information structurelle de ces variables. On va s'intéresser ici au cas général où ces structures sont des familles de Moore particulières et où on cherche à obtenir une famille consensus du même type. Ce problème peut se ramener à l'étude de l'ajustement possible d'une relation particulière R , contenant l'information commune au profil, à une classification $\mathcal{C}$. Nous avons démontré dans un article précédent

¹Computer Science Department, Université de Nicosie, 46 Makedonitissas Av., 1700 Nicosie, Chypre, domenach.f@unic.ac.cy

[Domenach, Leclerc, 2004(b)] que si un tel ajustement existe, satisfaisant deux conditions particulières liées aux conditions d'Adams, alors il est unique. Malheureusement, une telle classification n'existe pas pour toute relation R . L'existence dans le cas hiérarchique a été démontrée par [Adams III, 1986] en considérant une fonction consensus unanime.

Nous allons présenter ici une situation plus générale en considérant la fonction consensus obtenue avec les emboîtements présents dans au moins p familles du profil. Cette fonction consensus, avec les caractérisations de familles de Moore particulières obtenues dans [Domenach, Leclerc, 2007], va nous permettre de redémontrer le théorème d'Adams, et prouver l'existence d'une telle classification pour des familles de Moore emboîtées, des arbres de parties, et des géométries convexes avec la fonction de consensus unanime, et pour des familles de Moore distributives ou topologiques dans le cas général.

Cet article est structuré de la façon suivante : nous rappellerons d'abord dans la Section 2 les principales définitions concernant les familles de Moore et leurs différents cryptomorphismes, ainsi que différents types de familles de Moore qu'on retrouve en classification. La Section 3 présente le problème général d'aggrégation d'un profil de classifications avec des règles de quotas, et la Section 4 introduit deux généralisations possibles de telles règles. Les résultats principaux sont démontrés dans la Section 5, où nous définissons une fonction consensus d'emboîtements fréquents, et où nous caractérisons cette fonction consensus pour des hiérarchies, des arbres de parties, des familles de Moore distributives et des géométries convexes.

2. DÉFINITIONS

Une classification peut être considérée comme une famille $\mathcal{C}$ de sous-ensembles (classes) sur S . Nous allons ici évoquer un type de familles permettant de représenter des classifications, les familles de Moore.

2.1. CRYPTOMORPHISMES DE FAMILLES DE MOORE

Dans la suite de cet article, nous considérons un ensemble S fini, et $\mathcal{P}(S)$ l'ensemble de tous les sous-ensembles de S . Une famille $\mathcal{F}$ est une *famille de Moore* sur S (ou encore *système de fermeture*) si elle satisfait les deux propriétés suivantes: (i) $S \in \mathcal{F}$ et (ii) pour tout $A, B \in \mathcal{F}$, $A \cap B \in \mathcal{F}$. L'ensemble $\mathfrak{M}$ de toutes les familles de Moore sur S est aussi une famille de Moore, sur $\mathcal{P}(S)$. On peut associer à toute famille $\mathcal{C}$ de sous-ensembles de S la famille de Moore $\mu(\mathcal{C})$ définie par $\mu(\mathcal{C}) = \{\cap \mathcal{C}' : \mathcal{C}' \subseteq \mathcal{C}\} \cup \{S\}$.

Il existe de nombreux cryptomorphismes équivalents aux familles de Moore (c.f. [Caspard, Monjardet, 2003] pour une synthèse). Parmi ceux-ci, on peut distinguer les fermetures φ , applications sur $\mathcal{P}(S)$ *isotones* (pour tous $A, B \subseteq S, A \subseteq B \Rightarrow \varphi(A) \subseteq \varphi(B)$), *extensives* (pour tout $A \subseteq S, A \subseteq \varphi(A)$) et *idempotentes* (pour tout $A \subseteq S, \varphi(A) = \varphi(\varphi(A))$). L'image $\varphi(\mathcal{P}(S))$ est alors une famille de Moore, et, réciproquement, on définit une fermeture par $\varphi(A) = \cap \{F \in \mathcal{F} : A \subseteq F\}$ pour tout $A \subseteq S$. Il existe une autre caractérisation, où une application extensive φ sur $\mathcal{P}(S)$ est une fermeture si et seulement si φ vérifie la propriété d'*indépendance de chemin* [Plott, 1973] : $\varphi(A \cup B) = \varphi(\varphi(A) \cup \varphi(B))$

Un autre cryptomorphisme des familles de Moore, populaire en base de données relationnelles où ils trouvent leurs origines, sont les systèmes implicatifs complets. Un *système implicatif complet* I (ou $\rightarrow_I$, ou simplement $\rightarrow$) est une relation binaire sur $\mathcal{P}(S)$ satisfaisant les conditions suivantes, pour tous $A, B, C, D \subseteq S$:

- (I1) $B \subseteq A$ implique $A \rightarrow B$;
- (I2) $A \rightarrow B$ et $B \rightarrow C$ impliquent $A \rightarrow C$;
- (I3) $A \rightarrow B$ et $C \rightarrow D$ impliquent $A \cup C \rightarrow B \cup D$.

L'équivalence entre systèmes implicatifs complets et fermetures (et donc avec les familles de Moore) est due à [Armstrong, 1974] avec $A \rightarrow B \iff B \subseteq \varphi(A)$. La relation d'implication correspond à l'idée que la partie B est systématiquement associée à la partie A , en ce sens que A implique B si toute classe contenant A contient aussi B .

2.2. FAMILLES DE MOORE ET CLASSIFICATION

Il existe de nombreux types de familles de Moore, parmi lesquels nous allons distinguer par la suite les familles suivantes :

DÉFINITION 1. Soit $\mathcal{F}$ une famille de Moore sur S , et φ sa fermeture associée.

- $\mathcal{F}$ est emboîtée si $F, F' \in \mathcal{F}$ impliquent $F \cap F' \in \{F, F'\}$;
- $\mathcal{F}$ est un arbre de parties si $F, F' \in \mathcal{F}$ impliquent $F \cap F' \in \{\emptyset, F, F'\}$;
- $\mathcal{F}$ est hiérarchique si elle est un arbre de parties et si $\{s\} \in \mathcal{F}$ pour tout $s \in S$;
- $\mathcal{F}$ est distributive si $F, F' \in \mathcal{F}$ impliquent $F \cup F' \in \mathcal{F}$;
- $\mathcal{F}$ est une topologie si elle est distributive et si $\emptyset \in \mathcal{F}$;
- $\mathcal{F}$ 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 $\varphi(B) = F$;

On retrouve ces différents types de familles de Moore associés à différents schémas classificatoires. Par exemple, les arbres de classification, avec l'addition de l'ensemble vide, peuvent être vus comme des familles de Moore hiérarchiques. Ou encore, si on considère une variable v prenant ses valeurs dans le co-domaine D , i.e. $v(s) \in D$ pour tous $s \in S$, si D est un ordre total alors les valeurs observées induisent naturellement une famille de Moore emboîtée, plus précisément la collection de sous-ensembles $\{s \in S : v(s) \leq \alpha\}, \alpha \in D$. On peut trouver dans la table 1 quelques exemples de familles de Moore induites par différents schémas classificatoires [Domenach, Leclerc, 2004(b)].

Type de variable	Structure de D	Sous-ensembles de S	Type de famille de Moore
Numerical, ordinal	Ordre total	$\{s \in S : v(s) \leq \alpha\}, \alpha \in D$	Emboîté
		Intervalles de D	Géométrie convexe
Nominal	Ensemble fini $\{v_1, \dots, v_m\}$	$\{s \in S : v(s) = v_i\}$	Arbre de parties
Taxonomie	Arbre	$\{s \in S : v(s) \leq \alpha\}, \alpha \in D$	Arbre de parties
Evaluation multicritère	Produit d'ordres totaux	$\{s \in S : v(s) \leq \alpha\}, \alpha \in D$	Distributif

Table 1. Familles de Moore induites

3. CONSENSUS ET RÈGLES DE QUOTA

Dans cette section, nous allons considérer le problème consistant à agréger un profil de familles de Moore en une famille de Moore consensus. On peut se référer, parmi d'autres, à [Barthélemy, Janowitz, 1991 ; Leclerc, 1994] pour des résultats généraux, particulièrement portant sur les médianes.

Soit $\mathcal{F}^* = (\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_k) \in \mathfrak{M}^k$ un profil de classifications où toutes les familles de Moore $\mathcal{F}_i$ sont d'un type particulier (emboîté, hiérarchique, ...). Notre but est d'agréger $\mathcal{F}^*$ en une unique famille $\mathcal{F}$ du même type. Soit $K = \{1, 2, \dots, k\}$. On associe au profil $\mathcal{F}^*$ un indice $g_{\mathcal{F}^*}$ sur $\mathcal{P}(S)$ par, pour tout $A \subseteq S$,

$$g_{\mathcal{F}^*}(A) = |\{i \in K : A \in \mathcal{F}_i\}|$$

On associe la fonction consensus $F_{(p)} : \mathfrak{M}^k \rightarrow \mathfrak{M}$ à l'indice $g_{\mathcal{F}^*}$ pour tout entier $p \in K$. Un sous-ensemble A de S est p -fréquent si $g_{\mathcal{F}^*}(A) \geq p$, et le *consensus p -fréquent* de $\mathcal{F}^*$, noté $F_{(p)}(\mathcal{F}^*)$, est la famille de tous les ensembles p -fréquents. Comme dans le cas général $F_{(p)}(\mathcal{F}^*)$ n'est pas une famille de Moore, on va considérer le plus petit élément de $\mathfrak{M}$ le contenant, i.e. la fermeture de $F_{(p)}(\mathcal{F}^*)$ dans $\mathfrak{M}$. Parmi tous les consensus p -fréquents, on peut distinguer $F_{(k)}(\mathcal{F}^*) = \bigcap_{i \in K} \mathcal{F}_i$ l'ensemble des classes communes à toutes les familles du profil, i.e. la fonction de consensus unanime, et le consensus majoritaire $F_{(p)}$ où $p = \lceil \frac{k+1}{2} \rceil$ est le plus petit entier supérieur à $\frac{k}{2}$.

Notre fonction consensus $F_{(p)}$ est une règle de quota, cas particulier des fonctions consensus de fédération. Une *fédération* sur K est une famille $\mathcal{K}$ de sous-ensembles de K étant monotone: pour tout $L \in \mathcal{K}, L \subseteq L' \Rightarrow L' \in \mathcal{K}$. Alors la fonction consensus de fédération $c_{\mathcal{K}}$ sur $\mathfrak{M}$ est associée à $\mathcal{K}$ de la façon suivante :

$$c_{\mathcal{K}}(\mathcal{F}^*) = \bigvee_{L \in \mathcal{K}} \bigcap_{i \in L} \mathcal{F}_i$$

Les *règles de quotas* peuvent être vues comme des cas particuliers de telles fonctions, en prenant $\mathcal{K} = \{L \subseteq K : |L| \geq p\}$ pour un nombre p fixé, $0 \leq p \leq k$. De telles règles possèdent des propriétés intéressantes [Day et McMorris, 2003], par exemple :

Unanimité: pour tout $\mathcal{F} \in \mathfrak{M}$, $F_{(p)}(\mathcal{F}, \mathcal{F}, \dots, \mathcal{F}) = \mathcal{F}$;
Isotonie: pour tout $\mathcal{F}^* = (\mathcal{F}_1, \dots, \mathcal{F}_k)$ et $\mathcal{F}'^* = (\mathcal{F}'_1, \dots, \mathcal{F}'_k)$ profils sur $\mathfrak{M}^k$, $\mathcal{F}_i \subseteq \mathcal{F}'_i$ pour tout $i \in \mathcal{K}$ implique $F_{(p)}(\mathcal{F}^*) \subseteq F_{(p)}(\mathcal{F}'^*)$.

4. CONSENSUS FRÉQUENTS

En fait, comme il a été signalé dans la littérature, $F_{(p)}$ ne capture que peu d'informations communes aux différentes familles $\mathcal{F}_i$. Considérons l'exemple de la Figure 1 adaptée de [Adams III, 1972] où on a deux hiérarchies (moins l'ensemble vide) $\mathcal{H}_1$ et $\mathcal{H}_2$ pour lesquelles on cherche à extraire l'information commune. Si on considère simplement $F_{(2)}$, i.e. la fonction de consensus unanime, on va obtenir la hiérarchie $\mathcal{H}$ de la Figure 2, pauvre en information. En fait, $\mathcal{H}_1$ et $\mathcal{H}_2$ ont une riche structure commune, visualisée par $\mathcal{H}'$ (consensus d'Adams).

FIGURE 1. Deux hiérarchies sur $S = \{1, \dots, 8\}$

FIGURE 2. Hiérarchies obtenues en utilisant $F_{(2)}$ ($\mathcal{H}$) et $E_{(2)}$ ($\mathcal{H}'$)

Une des manières d'enrichir la famille consensus est de relacher la condition faisant qu'on ne considère que la présence ou l'absence de classes (ensembles fermés) communes aux familles du profil. Nous allons présenter deux possibles extensions à ce modèle permettant de ne plus considérer seulement les classes des familles, mais l'information commune: la première en considérant un profil de familles de Sperner (les familles de Moore et les profils de familles de Sperner étant en correspondance [Caspard et Monjardet, 2003; Demetrovics et Hua, 1991]). La deuxième, les relations d'emboîtement, est basée sur l'idée d'ensembles plus généraux que d'autres

par rapport à la famille considérée, et prend son origine dans [Adams III, 1986 ; Domenach, Leclerc, 2004(a)].

4.1. FAMILLES DE SPERNER

[Leclerc, 2007] considère le cas où $\mathcal{D}^* = (\mathcal{D}_1, \dots, \mathcal{D}_k)$ est un profil de familles de Sperner. Une famille de Sperner $\mathcal{D}$ est une famille de sous-ensembles de S deux à deux inclusion incomparables. On définit alors l'*indice de groupement* $g_{\mathcal{D}^*}$ par :

$$g_{\mathcal{D}^*}(A) = |\{i \in K : A \subseteq C \text{ pour au moins un ensemble } C \text{ de } \mathcal{D}_i\}|$$

et la fonction consensus $S_{(p)}(\mathcal{D}^*)$ comme la famille de Sperner de tous les groupements p -fréquents maximaux.

Ce cas modélise entre autres l'idée *itemsets fréquents* [Agrawal, Imielinski, Swami, 1993] chère à l'analyse des données lors de la recherche de règles d'association. Étant donné une base de données, on peut considérer chaque transaction comme une famille de Sperner réduite à une seule classe.

4.2. RELATIONS D'EMBOÎTEMENT

Une autre façon de généraliser cette règle de quotas F_p est de ne plus considérer les classes communes à p familles du profil, mais plutôt l'information commune. En reprenant l'exemple de la Figure 1, on peut remarquer que, dans les deux hiérarchies, pour les singletons $\{1\}, \{2\}$ et $\{3\}$, $\{2\}$ et $\{3\}$ sont plus proches. On souhaite donc retrouver dans la hiérarchie consensus une classe contenant $\{2\}$ et $\{3\}$ mais pas $\{1\}$.

Pour modéliser cette idée, considérons la relation binaire $\Subset$ introduite dans [Domenach, Leclerc, 2004(a)]. Cette *relation d'emboîtement* peut être définie axiomatiquement de la façon suivante :

- (O1) $A \Subset B$ implique $A \subset B$;
- (O2) $A \subset B \subset C$ implique $[A \Subset C \iff [A \Subset B \text{ ou } B \Subset C]]$;
- (O3) $A \Subset A \cup B$ implique $A \cap B \Subset B$.

Les relations d'emboîtement sont cryptomorphiques des familles de Moore et des fermetures. On peut associer à $\Subset$ une fermeture φ par $A \Subset B \iff A \subset B$ et $\varphi(A) \subset \varphi(B)$ pour tous $A, B \subseteq S$. Les relations d'emboîtement correspondent ainsi à l'idée intuitive que deux ensembles inclus l'un dans l'autre ont des fermetures différentes, i.e. que si A est emboîté dans B alors la partie B est plus générale que la partie A .

On peut définir de la même manière une relation d'emboîtement $\Subset_{\mathcal{C}}$ pour toute famille $\mathcal{C}$ de sous-ensembles de S ($\mathcal{C}$ n'étant pas forcément une famille de Moore) par $A \Subset_{\mathcal{C}} B \iff A \subset B$ et il existe $C \in \mathcal{C}$ tel que $A \subseteq C$ et $B \not\subseteq C$. De même que pour les implications, la relation d'emboîtement associée à un ensemble de classes quelconques $\mathcal{C}$ ne change pas lorsqu'on remplace $\mathcal{C}$ par la famille de Moore $\mu(\mathcal{C})$. On peut donc parler indifféremment de relation d'emboîtement sur $\mathcal{C}$ ou sur $\mu(\mathcal{C})$:

PROPOSITION 1. *Toute famille $\mathcal{C}$ de sous-ensembles de S a la même relation d'emboîtement que sa famille de Moore associée $\mu(\mathcal{C})$.*

Preuve. Soit φ la fermeture associée à $\mu(\mathcal{C})$. Soient $A, B \subseteq S$ tels que $A \in_c B$, i.e. il existe $C \in \mathcal{C}$ tel que $A \subseteq C$ et $B \not\subseteq C$. Si $\varphi(A) = \varphi(B)$, alors vu que C est un ensemble fermé, $\varphi(A) \subseteq C$ d'où $B \subseteq \varphi(B) = \varphi(A) \subseteq C$, contradiction. Donc $A \in_{\mu(\mathcal{C})} B$.

Inversement, soient $A, B \subseteq S$ tels que $A \notin_c B$, i.e. pour tout $C \in \mathcal{C}$, $A \subseteq C \Rightarrow B \subseteq C$. On a $A \subset B$ d'où $\varphi(A) \subseteq \varphi(B)$. Par définition de $\mu(\mathcal{C})$, il existe une sous-famille $\mathcal{C}_1$ de $\mathcal{S}$ telle que $\varphi(A) = \cap \mathcal{C}_1$, ce qui implique que pour tout $D \in \mathcal{C}_1$, $A \subseteq D$ d'où $B \subseteq D \Rightarrow B \subseteq \cap \mathcal{C}_1 = \varphi(A)$. Donc $A \subset B \subseteq \varphi(A) \subseteq \varphi(B)$ c'est à dire $\varphi(A) = \varphi(B)$.

La proposition suivante, qui se démontre facilement grâce à la définition de $\in$, nous sera utile pour les démonstrations de la section suivante :

PROPOSITION 2. *Pour tout $A \subset B \subseteq S$, on a les équivalences suivantes :*

$$A \in B \iff A \in \varphi(B) \iff \varphi(A) \in \varphi(B) \iff \varphi(A) \in B$$

5. EMBOÎTEMENTS FRÉQUENTS

Dans un article précédent [Domenach, Leclerc, 2004(b)], pour une relation binaire $R \subseteq (\mathcal{P}(S))^2$, nous avons montré qu'il ne pouvait exister qu'une seule classification $\mathcal{C}$ satisfaisant deux conditions liées à R et aux relations d'emboîtement, ces conditions nous assurant un ajustement possible de la relation d'emboîtement de $\mathcal{C}$ à R . Le problème de l'existence d'une telle classification restait ouvert. Adams [Adams III, 1986] a démontré l'existence en considérant des hiérarchies et une règle d'unanimité. Nous allons redémontrer ici le théorème d'Adams, et prouver, dans le cas de familles de Moore emboîtées, arbres de parties, distributives, topologiques et des géométries convexes, dans quelles conditions un tel ajustement est possible.

5.1. DÉFINITION

Soit $\mathcal{F}^* = (\mathcal{F}_1, \dots, \mathcal{F}_k) \in \mathfrak{M}^k$ un profil de classifications, avec $\in_i$ la relation d'emboîtement associée à la famille de Moore $\mathcal{F}_i$, pour tout $i \in K$, et $E_{(p)}$ le groupement p -fréquent défini par $E_{(p)}(\mathcal{F}^*) = \bigcup_{J \subseteq K, |J| \geq p} \bigcap_{i \in J} \in_i$. On peut considérer $E_{(p)}$ comme l'ensemble de toutes les paires $(A, B) \in \mathcal{P}(S) \times \mathcal{P}(S)$ qui appartiennent à au moins p relations d'emboîtement $\in_i$. Dans le cas général, $E_{(p)}$ n'est pas une relation d'emboîtement, donc lorsqu'on cherche à agréger un profil $\mathcal{F}^*$ en une famille de Moore unique $\mathcal{F}^c$, ayant $\in^c$ comme relation d'emboîtement, on ne peut demander l'égalité $\in^c = E_{(p)}(\mathcal{F}^*)$.

Considérons les deux conditions suivantes :

(R1)	$E_{(p)} \subseteq \in^c$	<i>Emboîtements préservés</i>
(R2)	Pour tout $F, F' \in \mathcal{F}^c$, $F \subset F'$ implique $(F, F') \in E_{(p)}$	<i>Emboîtements confirmés</i>

Ces deux conditions prennent leurs origines dans [Adams III, 1986], où l'inclusion (R1) correspond aux *emboîtements préservés*, i.e. les emboîtements présents dans au moins p familles du profil. La condition (R2) d'*emboîtements confirmés* peut être vue comme la réciproque partielle de (R1), où seul les ensembles fermés de $\mathcal{F}^c$ sont emboîtés dans au moins p famille $\mathcal{F}_i$. En notant par $\mathcal{M}_{\mathcal{F}^c}$ l'ensemble des éléments

$\cap$ -irréductibles de $\mathcal{F}^c$, i.e. ne s'obtenant pas comme intersection d'autres éléments de $\mathcal{F}^c$ et par M^+ l'élément de $\mathcal{F}^c$ couvrant M pour tout $M \in \mathcal{M}_{\mathcal{F}^c}$, la condition (R2) est équivalente à la condition suivante [Leclerc, 2004] :

(R2') Pour tout $M \in \mathcal{M}_{\mathcal{F}^c}$, $M \in M^+$

Par la suite, nous écrirons indifféremment $(A, B) \in E_{(p)}$ ou $(A, B) \in E_{(p)}(\mathcal{F}^*)$ lorsqu'il n'y a pas de confusion possible. Étant donné que les implications sont plus populaires que les relations d'emboîtement, on peut formuler deux conditions (RI1) et (RI2) équivalentes à (R1) et (R2) de la façon suivante [Leclerc, 2007] :

(RI1) Pour tous $A, B \subseteq S$, $A \rightarrow_c B$ implique $|\{i \in K : A \rightarrow_i B\}| \geq k - p$

(RI2) Pour tout $F, F' \in \mathcal{F}^c$, $F \subset F'$ implique $|\{i \in K : F \rightarrow_i F'\}| < k - p$

L'unicité d'un tel groupement a été prouvée dans [Domenach, Leclerc, 2004(b)] dans le cas plus général d'une relation binaire R sur $\mathcal{P}(S)$, ayant comme unique condition que $(A, B) \in R$ implique $A \subset B$. Si on considère une famille de Moore $\mathcal{F}$, avec $\subseteq$ la relation d'emboîtement associée, les deux propriétés (R1) et (R2) peuvent être vues comme cas particuliers de :

(AR1) $R \subseteq \subseteq$;

(AR2) for all $F, G \in \mathcal{F}$, $F \subset G$ implique $(F, G) \in R$.

THÉORÈME 1. Soient $\mathcal{F}'$ et $\mathcal{F}$ deux familles de Moore sur $\mathcal{P}(S)$.

Si $\mathcal{F}$ et $\mathcal{F}'$ satisfont les conditions (AR1) and (AR2), alors $\mathcal{F} = \mathcal{F}'$.

L'existence d'une telle relation binaire restait un problème dans le cas général décrit dans [Domenach, Leclerc, 2004(b)]. Toutefois nous avons ici, par construction, une relation consensus. Le problème devient donc de savoir si ce consensus conserve le type de famille de Moore du profil, i.e. dans quelles conditions nous avons une famille hiérarchique, par exemple, lorsque le profil à agréger est composé de familles du même type. La section suivante donne des résultats pour quelques familles spécifiques, en nous basant sur des caractérisations par les relations d'emboîtement de propriétés des familles de Moore faites dans [Domenach, Leclerc, 2007].

5.2. HIÉRARCHIES ET ARBRES DE PARTIES

Le fait que tous les singletons d'une famille de Moore soient des ensembles fermés se caractérise facilement grâce aux relations d'emboîtement de la façon suivante :

(OS) $A \not\in \{\emptyset, \{s\}\}$ implique $\{s\} \subseteq A \cup \{s\}$, pour tout $s \in S$

PROPOSITION 3. Pour tout profil $\mathcal{F}^* = (\mathcal{F}_1, \dots, \mathcal{F}_k)$ de familles de Moore sur S , et pour tout $p \in K$, on a l'équivalence suivante :

$E_{(p)}$ satisfait (OS) si et seulement si $\subseteq^c$ satisfait (OS).

Preuve. La condition nécessaire est triviale avec (R1). Pour la condition suffisante, soit $A \not\in \{\emptyset, \{s\}\}$ un sous-ensemble de S . $A \subseteq^c A \cup \{s\}$ implique $\varphi(A) \subset \varphi(A \cup \{s\})$, donc, par (R2), $(\varphi(A), \varphi(A \cup \{s\})) \in E_{(p)}$, i.e. $(A, A \cup \{s\}) \in E_{(p)}$.

La propriété caractéristique des hiérarchies que toute paire d'ensembles fermés sont ou d'intersection vide, ou inclus l'un dans l'autre, se traduit pour la relation d'emboîtement associée par la propriété (OT) :

(OT) $A \subseteq C$ et $B \subseteq C$ impliquent $A \cup B \subseteq C$ ou $A \cap B = \emptyset$, pour tous $A, B, C \subseteq S$

PROPOSITION 4. *Pour tout profil $\mathcal{F}^* = (\mathcal{F}_1, \dots, \mathcal{F}_k)$ de familles de Moore sur S , on a l'équivalence suivante :*

$E_{(k)}$ satisfait (OT) si et seulement si $\mathbb{E}^c$ satisfait (OT).

Preuve. Considérons deux sous-ensembles A et B de S tels que $A \cap B \neq \emptyset$. Pour la condition nécessaire, soit $C \subseteq S$ tel que $A \mathbb{E}^c C$ et $B \mathbb{E}^c C$. Par définition, $\varphi(A) \subset \varphi(C)$ et $\varphi(B) \subset \varphi(C)$, donc, par (R2), $(\varphi(A), \varphi(C)) \in E_{(k)}$ et $(\varphi(B), \varphi(C)) \in E_{(k)}$. D'où $(\varphi(A) \cup \varphi(B), \varphi(C)) \in E_{(k)}$, et par la propriété d'indépendance de chemin, $(\varphi(A \cup B), \varphi(C)) \in E_{(k)}$, et donc, par (R1), $A \cup B \mathbb{E}^c C$.

Pour la condition suffisante, soit $C \subseteq S$ tel que $(A, C) \in E_{(k)}$ et $(B, C) \in E_{(k)}$. Par (R1), $A \mathbb{E}^c C$ et $B \mathbb{E}^c C$ d'où $A \cup B \mathbb{E}^c C$ implique $\varphi(A \cup B) \subset \varphi(C)$ i.e. $(\varphi(A \cup B), \varphi(C)) \in E_{(k)}$ et donc $(A \cup B, C) \in E_{(k)}$.

En fait, c'est la condition nécessaire de la Proposition 4 qui oblige à considérer la relation de consensus unanime $E_{(k)}(\mathcal{F}^*) = \bigcap_{i \in K} \mathbb{E}_i$ à la place d'un plus général groupement p -fréquent $E_{(p)}$. En effet, lorsqu'on considère la conjonction des deux conditions $(A, C) \in E_{(p)}$ et $(B, C) \in E_{(p)}$, A est emboîté dans C dans au moins p familles, de même pour B et C , mais rien ne garantit qu'il va exister au moins p familles pour lesquels ces deux conditions sont vraies en même temps. La seule valeur de p pour laquelle la condition (OT) se retrouve dans la famille consensus est donc pour $p = k$.

Les deux propositions précédentes, caractérisant les familles de Moore hiérarchiques, nous permettent de redémontrer facilement le théorème de consensus de hiérarchies d'Adams [Adams III, 1986] :

COROLLAIRE 1 (Théorème d'Adams). *Soit $\mathcal{H}^*$ un profil de familles de Moore hiérarchiques sur S , et $E_{(k)}(\mathcal{H}^*)$ le groupement p -fréquent unanime.*

Alors, si $\mathbb{E}^c$ est la relation d'emboîtement satisfaisant les conditions (R1) et (R2) associée à $E_{(k)}$, la famille de Moore consensus $\mathcal{H}^c$ correspondant à $\mathbb{E}^c$ est encore hiérarchique.

Preuve. L'existence est conséquence des Propositions 3 et 4, et l'unicité provient de [Domenach et Leclerc, 2004(b)].

La preuve pour la proposition suivante est similaire à celle de la Proposition 4, concerne les familles de Moore emboîtées :

(ON) $A \mathbb{E} C$ et $B \mathbb{E} C$ impliquent $A \cup B \mathbb{E} C$, pour tous $A, B, C \subseteq S$

PROPOSITION 5. *Pour tout profil $\mathcal{F}^* = (\mathcal{F}_1, \dots, \mathcal{F}_k)$ de familles de Moore sur S , on a l'équivalence suivante :*

$E_{(k)}$ satisfait (ON) si et seulement si $\mathbb{E}^c$ satisfait (ON).

COROLLAIRE 2. *Soit $\mathcal{N}^*$ un profil de familles de Moore emboîtées sur S , et $E_{(k)}(\mathcal{N}^*)$ le groupement p -fréquent unanime.*

Alors, si $\mathbb{E}^c$ est la relation d'emboîtement satisfaisant les conditions (R1) et (R2) associée à $E_{(k)}$, la famille de Moore consensus $\mathcal{N}^c$ correspondant à $\mathbb{E}^c$ est encore emboîtée.

Une relation binaire $\Subset$ est la relation d'emboîtement associée à un arbre de parties si et seulement si elle satisfait les propriétés (O1), (O2) et (OT), ou (O1), (O2), (OE) et (ON) [Domenach, Leclerc, 2007]. On a donc le corollaire suivant :

COROLLAIRE 3. *Soit $\mathcal{A}^*$ un profil d'arbres de parties sur S , et $E_{(k)}(\mathcal{A}^*)$ le groupement p -fréquent unanime.*

Alors, si $\Subset^c$ est la relation d'emboîtement satisfaisant les conditions (R1) et (R2) associée à $E_{(k)}$, la famille de Moore consensus $\mathcal{S}^c$ correspondant à $\Subset^c$ est encore un arbre de parties.

5.3. FAMILLES DE MOORE DISTRIBUTIVES ET TOPOLOGIES

La propriété (OD) suivante particularise les familles de Moore distributives:

$$(OD) \quad \text{pour tous } s \in S, A \subseteq S, [\{a\} \Subset \{a, s\} \text{ pour tout } a \in A] \iff [A \Subset A \cup \{s\}]$$

PROPOSITION 6. *Pour tout profil $\mathcal{F}^* = (\mathcal{F}_1, \dots, \mathcal{F}_k)$ de familles de Moore sur S , et pour tout $p \in K$, on a l'équivalence suivante :*

$$E_{(p)} \text{ satisfait (OD) si et seulement si } \Subset^c \text{ satisfait (OD)}.$$

Preuve. Montrons que $E_{(p)}$ satisfait (OD) implique $\Subset^c$ satisfait (OD). Soient $A \subseteq S$ et $s \in S$ tels que $A \Subset A \cup \{s\}$. Ce qui est équivalent à $\varphi(A) \subset \varphi(A \cup \{s\})$ et implique par (R2) $(\varphi(A), \varphi(A \cup \{s\})) \in E_{(p)}$, i.e. $(A, A \cup \{s\}) \in E_{(p)} \iff (\{a\}, \{a, s\}) \in E_{(p)}$ pour tout $a \in A$ d'où $\{a\} \Subset^c \{a, s\}$. L'implication réciproque se démontre de manière analogue.

Inversement, soient $A \subseteq S$ et $s \in S$ tels que $(A, A \cup \{s\}) \in E_{(p)}$. Ce qui implique par (R1) $A \Subset A \cup \{s\} \iff \{a\} \Subset^c \{a, s\}$ pour tout $a \in A$ et donc $\varphi(\{a\}) \subset \varphi(\{a \cup s\}) \Rightarrow (\varphi(\{a\}), \varphi(\{a \cup s\})) \in E_{(p)}$. L'implication réciproque est similaire.

COROLLAIRE 4. *Soit $\mathcal{D}^*$ un profil de familles de Moore distributives sur S , et $E_{(p)}(\mathcal{D}^*)$ le groupement p -fréquent.*

Alors, si $\Subset^c$ est la relation d'emboîtement satisfaisant les conditions (R1) et (R2) associée à $E_{(p)}$, la famille de Moore consensus $\mathcal{D}^c$ correspondant à $\Subset^c$ est encore distributive.

Preuve. De façon similaire au corollaire 1, l'existence provient de la Proposition 6, et l'unicité de [Domenach, Leclerc, 2004(b)].

La propriété (OE) suivante est caractéristique des familles de Moore ayant l'ensemble vide comme ensemble fermé, i.e. comme infimum du treillis :

$$(OE) \quad \emptyset \Subset \{s\} \text{ pour tout } s \in S$$

PROPOSITION 7. *Pour tout profil $\mathcal{F}^* = (\mathcal{F}_1, \dots, \mathcal{F}_k)$ de familles de Moore sur S , et pour tout $p \in K$, on a l'équivalence suivante :*

$$E_{(p)} \text{ satisfait (OE) si et seulement si } \Subset^c \text{ satisfait (OE)}.$$

Preuve. La condition nécessaire est triviale en utilisant (R1). Pour la condition suffisante, on a $\emptyset \in^c \{s\}$ pour tout $s \in S$, i.e. $\varphi(\emptyset) = \emptyset \subset \varphi(\{s\})$. D'où par (R2) il y a au moins p relations d'emboîtement $\in_i$ tels que $\emptyset \in_i \varphi(\{s\})$ donc $\emptyset \in_i \{s\}$.

Les familles de Moore topologiques étant caractérisées par la conjonction des conditions (O1), (O2), (O3), (OD) et (OE), on a le corollaire suivant :

COROLLAIRE 5. *Soit $\mathcal{T}^*$ un profil de familles de Moore topologiques sur S , et $E_{(p)}(\mathcal{T}^*)$ le groupement p -fréquent.*

Alors, si $\in^c$ est la relation d'emboîtement satisfaisant les conditions (R1) et (R2) associée à $E_{(p)}$, la famille de Moore consensus $\mathcal{T}^c$ correspondant à $\in^c$ est encore topologique.

5.4. GÉOMÉTRIES CONVEXES

La propriété caractéristique des géométries convexes trouve aussi son équivalent grâce aux relations d'emboîtement :

(OC) si $A \subseteq C$ et $B \subseteq C$, alors $A \cap B \in C$ implique $A \in C$ ou $B \in C$, pour tous $A, B, C \subseteq S$

PROPOSITION 8. *Pour tout profil $\mathcal{F}^* = (\mathcal{F}_1, \dots, \mathcal{F}_k)$ de familles de Moore sur S , on a l'équivalence suivante :*

$E_{(k)}$ satisfait (OC) si et seulement si $\in^c$ satisfait (OC).

Preuve. Soient $A, B, C \subseteq S$ tels que $A \subseteq C$ et $B \subseteq C$. Si $A \cap B \in^c C$ alors $\varphi(A \cap B) \subset \varphi(C)$ i.e., par (R2), $(\varphi(A \cap B), \varphi(C)) \in E_{(k)}$ et donc $(A \cap B, C) \in E_{(k)}$. Supposons que $A \notin^c C$, c'est-à-dire $(A, C) \notin E_{(k)}$, alors par hypothèse $(B, C) \in E_{(k)}$, donc $B \in^c C$. Inversement, $(A \cap B, C) \in E_{(k)}$ implique $A \cap B \in^c C$ d'où $A \in^c C$ ou $B \in^c C$, i.e. $\varphi(A) \subset \varphi(C)$ ou $\varphi(B) \subset \varphi(C)$. Par (R2), on a $(\varphi(A), \varphi(C)) \in E_{(k)}$ ou $(\varphi(B), \varphi(C)) \in E_{(k)}$ et donc $(A, C) \in E_{(k)}$ ou $(B, C) \in E_{(k)}$.

COROLLAIRE 6. *Soit $\mathcal{G}^*$ un profil de géométries convexes sur S , et $E_{(k)}(\mathcal{G}^*)$ le groupement p -fréquent unanime.*

Alors, si $\in^c$ est la relation d'emboîtement satisfaisant les conditions (R1) et (R2) associée à $E_{(k)}$, la famille de Moore consensus $\mathcal{G}^c$ correspondant à $\in^c$ est encore une géométrie convexe.

Preuve. De même que pour le Corollaire 1, on a l'existence comme conséquence de la Proposition 8, et l'unicité provient de [Domenach, Leclerc, 2004(b)].

6. CONCLUSION

Nous avons défini une classe de règles de consensus par emboîtements fréquents applicables pour différents profils $\mathcal{F}^*$ de familles de Moore, et nous avons montré sous quelles conditions on peut ajuster une classification consensus lorsqu'on a des familles de Moore emboîtées, hiérarchiques, arbres de parties, distributives, topologiques ou encore pour des géométries convexes. Il reste à généraliser ces résultats d'abord à d'autres types de familles de Moore (hiérarchies faibles

[Bandelt, Dress, 1989], hiérarchies k -faibles [Bandelt, Dress, 1994], pyramides [Diday, 1986], ...) mais aussi à considérer le cas où on cherche à agréger un profil de familles d'un certain type, par exemple des arbres de parties, en une famille consensus d'un type différent, par exemple hiérarchique. Étant donné la taille des relations d'emboîtement, un autre aspect à étudier est l'approche algorithmique pour obtenir un consensus par emboîtements fréquents.

Remerciements. L'auteur remercie le rapporteur anonyme pour ses commentaires et suggestions, notamment concernant diverses clarifications nécessaires.

BIBLIOGRAPHIE

- ADAMS III E.N. (1972), "Consensus Techniques and the Comparison of Taxonomic Trees", *Systematic Zoology* 21, p. 390-397.
- ADAMS III E.N. (1986), "N-trees as nestings: complexity, similarity and consensus", *Journal of Classification* 3, p. 299-317.
- AGRAWAL R., IMIELINSKI T., SWAMI A.N. (1993), "Mining Association Rules between Sets of Items in Large Databases", *SIGMOD* 22(2), p. 207-216.
- ARMSTRONG W.W. (1974), "Dependency structures of data base relationships", *Information Processing* 74, p. 580-583.
- BANDELT H.J., DRESS A.W.N. (1989), "Weak hierarchies associated with similarity measures: an additive clustering technique", *Bulletin Mathematical Biology* 51, p. 113-166.
- BANDELT H.J., DRESS A.W.N. (1994), "An order theoretic framework for overlapping clustering", *Discrete Mathematics* 136, p. 21-37.
- BARTHÉLEMY J.-P., JANOWITZ M.F. (1991), "A formal theory of consensus", *SIAM Journal of Discrete Mathematics* 4, p. 305-322.
- CASPARD N., MONJARDET B. (2003), "The lattices of Moore families and closure operators on a finite set: a survey", *Discrete Applied Mathematics* 127, p. 241-269.
- CROWN J.D., JANOWITZ M.F. (2003), "An injective set representation of closed systems of sets", *Bioconsensus: DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 61, p. 67-79.
- DAY W.H.E., MCMORRIS F.R. (2003), "Axiomatic Consensus Theory in Group Choice and Biomathematics", *SIAM Frontiers in Mathematics*, Philadelphia (PA).
- DEMETROVICS J., HUA N.S. (1991), "Databases, closure operations and Sperner families", (P. Frank *et al.* eds.), in *Extremal Problems for Finite Sets*, Budapest, János Bolyai Mathematical Society, p. 199-203.
- DIDAY E. (1986) DIDAY E., « Une représentation visuelle des classes empiétantes : les pyramides », *RAIRO APII* 20(5), p. 475-526.
- DOMENACH F., LECLERC B. (2004(a)), "Closure Systems, Implicational Systems, Overhanging Relations and the case of Hierarchical Classification", *Mathematical Social Sciences* 47, p. 349-366.
- DOMENACH F., LECLERC B. (2004(b)), "Consensus of classification systems, with Adams' results revisited", in Banks D., House L., McMorris F.R., Arabie P. and Gaul W. (eds), *Classification, Clustering and Data Mining Applications*, Berlin, Springer, p. 417-428.

- DOMENACH F., LECLERC B. (2007), "The structure of the overhanging relations associated with some types of closure systems", *Annals of Mathematics and Artificial Intelligence* 49, p. 37-149.
- LECLERC B., CUCUMEL G. (1987), "Consensus en classification: une revue bibliographique", *Mathématiques et Sciences humaines* 100, p. 109-128.
- LECLERC B. (1994), "Medians for weight metrics in the covering graphs of semilattices", *Discrete Applied Math.* 49, p. 281-297.
- LECLERC B. (2004), "On the consensus of closure systems", *Annales du LAMSADE* 3, p. 237-247.
- LECLERC B. (2007), "Consensus of classifications based on frequent groupings", (sous la direction de P. Brito, P. Bertrand, G. Cucumel et F. De Carvalho), in *Selected Contributions in Data Analysis and Classification*, Berlin, Springer, p. 317-324.
- PLOTT C.R. (1973), "Path independence, rationality and social choice", *Econometrica* 41(6), p. 1075-1091.