
Influence des classes intermédiaires dans les tests de classification

Impact of non-terminal classes on classification tests

Jérôme Euzenat

*INRIA Rhône-Alpes
655 avenue de l'Europe, 38330 Montbonnot Saint-Martin, France
Jerome.Euzenat@inrialpes.fr*

RÉSUMÉ. Dans le cadre d'une tâche de conception de hiérarchie, on mets en évidence l'influence des classes intermédiaires (ayant des sous-classes) sur le type de taxonomie obtenue (avec ou sans multi-spécialisation).

ABSTRACT. In the framework of taxonomic design, we provide evidence for the influence of non-terminal classes on the kind of resulting taxonomy (with or without multiple super-classes).

MOTS-CLÉS : construction de taxonomie — génie logiciel — protocole expérimental.

KEY WORDS : taxonomic design — software engineering — experimental protocol.

1. Objectif

Dans [Ouaggag 1997], Hamid Ouaggag et Robert Godin reprennent une expérience mise en œuvre dans [Dvorak 1994] demandant à un groupe d'individus de créer une classification à partir d'un ensemble de classes donné. Le résultat mis en évidence dans [Dvorak 1994] est que plus les classes se trouvent basses dans la classification, plus les sujets sont en désaccord sur leurs super-classes (par calcul du degré d'accord — c'est-à-dire la comparaison du nombre de classes ayant les mêmes super-classes).

Le but de [Ouaggag 1997] est de comparer les effets d'une classification avec de la multi-généralisation (le fait d'autoriser plusieurs super-classes immédiates pour une classe) par rapport à ceux d'une classification sans multi-généralisation. Nous appellerons la première activité (et le résultat de cette activité) multi-classification et la seconde (et son résultat) mono-classification. Ils posent donc la même question que dans [Dvorak 1994] à deux groupes témoins, l'un devant répondre par une multi-classification, l'autre par une mono-classification. Les classes choisies pour cette tâche sont les classes SmallTalk suivantes :

Array - Collection - Dictionary - ExtensibleCollection - IndexedCollection - Interval - LinkedList - OrderedCollection - OrderedList - SequenceableCollection - Set - SortedCollection - String

Le résultat obtenu est que la multi-classification est plus homogène que la mono-classification (en utilisant toujours la même mesure).

Cependant, il nous semble que la liste de classes utilisée introduit un biais dans l'expérience qui profite à la multi-classification. Le problème est que ces hiérarchies décrivent une classification en fonction de plusieurs hiérarchies indépendantes (donc plutôt multiples, ce que nous allons tester). Ceci pourrait expliquer que les sujets de l'expérience de [Ouaggag 1997] aient plus de facilité à faire de la multi-classification ce qui se traduit par une plus grande homogénéité des classifications produites. Nous relatons ici une expérience destinée à tester cette intuition.

Attention cependant, le test présenté ici n'est pas de même nature que celui de [Ouaggag 1997] : il ne prend pas en compte la cohésion entre les classifications. On ne peut donc conclure à l'inverse de leurs résultats mais juste suggérer une contre-expérience avec d'autres classes.

2. Méthodologie

Sujets de l'expérience : 13 membres d'une équipe de recherche partiellement liée aux objets (niveau 3e cycle). Les sujets sont donc plus qualifiés que ceux de [Ouaggag 1997] mais ils ne connaissent pas SmallTalk ni le contenu de la hiérarchie en question.

Principe : Les sujets vont avoir à réaliser deux exercices (dont le texte est donné ci-après) consistant à construire une classification. Ils sont libres d'utiliser mono- ou multi-classification. Le premier exercice concerne des classes de bas niveau où les sujets sont libres d'organiser leur classification, alors que dans le second exercice des classes intermédiaires sont ajoutées (on a alors les classes cités précédemment) et les sujets ne peuvent en introduire de nouvelles. Ainsi, on cherche à évaluer l'impact de la présence des classes intermédiaires dans le test. Si elle n'est pas significative, il n'y a alors aucun biais dans le test de [Ouaggag 1997].

Protocole : Le test est introduit par un bref rappel sur ce que peut être une classification, une présentation de mono- et multi-classification sur des exemples concernant des diplômes. Puis vient une brève présentation de ce que sont les classes du premier exemple (deux phrases par classe).

À la suite de quoi il est demandé aux sujets de tracer sur une feuille de papier (en dessous de la question) un graphe correspondant au premier exercice. Lorsqu'ils ont terminé ils doivent se lever et amener la feuille à l'instructeur sur laquelle celui-ci note le temps écoulé.

Quand tous les sujets ont terminé, les nouvelles classes sont présentées aussi brièvement qu'avant. Il est alors demandé de réaliser le second exercice dans les mêmes conditions que précédemment.

Modalités : Les sujets n'ont pas été récompensés (à l'exception d'un petit déjeuner) et il leur a été indiqué qu'ils ne seraient pas notés et qu'il n'y avait pas de

réponse unique à la question posée. Bien entendu ils ignoraient le but de l'expérience.

Problèmes possibles : Contrairement à une méthodologie classique de psychologie expérimentale les sujets sont en groupe (dans le cadre présent je pense que c'est acceptable) et l'expérience n'a pas été faite en double aveugle (les explications étaient données par un instructeur connaissant le but de l'expérience). Il n'y a pas de limitation de temps (mais sans doute une petite pression sur les sujets du fait qu'ils soient en groupe et qu'ils se lèvent lorsqu'ils ont terminé). Le fait qu'ils soient en groupe (et qu'ils puissent se rendre compte du temps mis par les autres) influe aussi certainement sur la différence entre la première et la seconde série de question. Ce problème influant principalement sur les temps de réponse, il n'en sera pas tenu compte dans l'analyse des résultats.

3. Les deux exercices (tels que présentés par écrit aux sujets)

Exercice 1 : voici un ensemble de structures de données (informatique).

On vous demande de les organiser dans une structure de spécialisation en introduisant d'autres classes si besoin.

Array - Dictionary (hash-table) - Interval - LinkedList - OrderedList - Set - String

Exercice 2 : faites l'exercice précédant sans introduire de nouvelles classes à l'aide des classes suivantes :

Array - Collection - Dictionary - ExtensibleCollection - IndexedCollection - Interval - LinkedList - OrderedCollection - OrderedList - SequenceableCollection - Set - SortedCollection - String

4. Résultats

Les résultats sont donnés dans les tableaux 1 et 2.

	Exercice 1	Exercice 2
Nombre de sujets ayant réalisés une mono-classification	7	0
Nombre de sujets ayant réalisés une multi-classification	6	13
Temps moyen en mono-classification	3'53	n.a.
Temps moyen en multi-classification	7'08	8'24

Table 2 : synthèse des résultats du tableau 1 (les moyennes temporelles ne sont pas significative pour les standards utilisés en psychologie).

	Exercice 1			Exercice 2			
	Temps	Nb cl.	Nb cl. M / cl.	Temps	Nb cl.	Nb cl. M	cl. M / cl.
1'13	9	0	n.a.	4'28	13	3	23%
2'55	13	1	8%	4'41	—	5	23%
3'05	8	0	n.a.	5'21	—	3	38%
3'46	9	1	11%	6'27	—	1	8%
4'08	7 (sic)	0	n.a.	8'21	—	4	31%
4'13	8	0	n.a.	8'28	—	5	38%
4'50	10	0	n.a.	8'39	—	2	15%
5'04	11	0	n.a.	9'10	—	4	31%
5'25	11	0	n.a.	9'14	—	6	46%
6'54	12	6	50%	9'15	—	2	15%
7'18	12	6	50%	9'26	—	2	15%
10'20	12	1	8%	11'32	—	1	8%

Table 1 (Profil temporel) : On y relève le temps passé (en mn), le nombre de classes dans la classification, le nombre de classes dans la classification ayant plusieurs super-classes (bien entendu l'exercice 2 impose le nombre de classes) et le pourcentage de tels classes par rapport au nombre total de classes (non applicable dans de nombreux cas de l'exercice 1). On peut noter que l'un des sujets du premier exercice n'a pas réalisé l'exercice puisqu'il ne cite pas toutes les classes.

5. Analyse

1) La question posée dans l'expérience initiale (exercice 2) est effectivement orientée vers la multi-classification (ou héritage multiple) alors qu'une organisation libre (exercice 1) va vers des hiérarchies plutôt mono-classifiées.

2) Ceci est encore renforcé par le comptage des classes puisque l'on peut considérer que beaucoup de réponses à la première question sont principalement en mono-classification et que les sujets ont ajouté un arc pour simplement se tirer d'une situation délicate (si l'on compte les résultats dont une unique classe a plusieurs super-classes comme mono-classifiés, on se trouve alors avec des résultats de seulement 3 multi-classifications dans le premier exercice contre 10 dans le second).

3) Lors de la première question on observe que les tenants de la multi-classification passent notablement plus de temps que ceux de la mono-classification.

4) Le second test prend plus de temps, même par rapport aux multi-classifications du premier. Cela peut être dû à plusieurs facteurs :

- les multi-classifications sont plus complexes que les mono-classifications ;
- les contraintes posées par le second exercice sont plus exigeantes, même pour ceux qui maîtrisent la multi-classification ;
- on peut aussi penser à un biais dans notre méthodologie, qui fait que certains s'apercevant que l'on avait attendu les « retardataires » lors du premier exercice ont pris plus de temps pour le second.

Conclusion

Le but de cette expérience était de tester la possibilité d'un biais en faveur de la multi-classification dans les classes utilisées dans [Ouaggag 1997]. Ce point nous semble acquis.

Il serait donc nécessaire de développer une expérience similaire à celle de [Ouaggag 1997] soit avec un ensemble de classes neutres (mais quelles sont elles?), soit en plaçant les utilisateurs en position de réels concepteurs, c'est-à-dire en les laissant libre de créer les classes intermédiaires eux-mêmes.

Remerciements

André Bisseret a bien voulu se pencher sur ma petite expérience et tenter de m'inculquer les principes du traitement statistique des données expérimentales.

Robert Godin a donné son accord à la présentation de ce poster (bien vite fait) dans la conférence même où il présente ses résultats. Je les en remercie.

Bibliographie

Joseph Dvorak, Conceptual entropy and its effects on class hierarchies, *IEEE Computer* 27(6):59-63, 1994

Hamid Ouaggag, Robert Godin, Comparaison empirique de l'effet sur l'entropie conceptuelle de l'héritage multiple versus l'héritage simple, Actes 4e conférence « Langages et modèles à objets », Roscoff (FR), 1997 à paraître