

Réseaux Bayésiens pour la Découverte de Stratégies d'Écriture Manuscrite chez l'Enfant en École Primaire

A Bayesian Network Model for Discovering Handwriting Strategies of Primary School Children

Ph. Leray¹, I. Zaarour^{1,2}, L. Heutte¹, B. El Eter³, J. Labiche¹, D. Mellier⁴

¹ Laboratoire PSI - FRE CNRS 2645, Université et INSA de Rouen, France

² Faculté de génie, branche I, Université Libanaise, Beyrouth, Liban

³ Laboratoire LPM, Université Libanaise, Beyrouth, Liban

⁴ Laboratoire PSY.CO (EA 1780), Université de Rouen, France

INSA Rouen / PSI, FRE CNRS 2645

BP 08 - Av. de l'Université, 76801 St-Etienne du Rouvray Cedex

Philippe.Leray@insa-rouen.fr

Résumé

Nous présentons dans cet article l'étude du suivi de l'évolution de l'écriture parmi des enfants de scolarité primaire. Nous proposons une méthode permettant de découvrir des groupes d'enfants partageant les mêmes stratégies d'écriture. A partir d'une acquisition en ligne de tests d'écriture et de dessins, nous modélisons ce problème grâce à un modèle graphique probabiliste, et plus précisément d'un réseau bayésien. Des expertises déterminent partiellement la structure du réseau dans lequel la stratégie d'écriture est représentée par une variable latente. L'architecture hiérarchique que nous proposons considère que chaque test est représenté par sa propre stratégie (locale) et qu'il en existe une globale prenant en compte chaque stratégie locale. L'algorithme EM nous permet d'estimer les paramètres de ce classifieur non supervisé qui relie les variables réellement mesurées (caractéristiques issues de l'acquisition d'écriture ou de dessins) et les stratégies locales et globale. Nous présentons ici des résultats préliminaires concernant la première étape de labélisation des stratégies découvertes par notre modèle et l'évolution temporelle des enfants dans ces stratégies.

Mots Clef

Stratégies d'écriture manuscrite, réseaux bayésiens, variables latentes, algorithme EM.

Abstract

In this paper, we present the results of a longitudinal study for the evolution follow-up in writing among typical pupils in primary education. We propose a method aimed at discovering groups of pupils sharing the same writing strategies. From the on-line acquisition of writing and drawing tests, this problem is modeled by means of a probabilistic graphical model, i.e. a bayesian network. Expert knowledge partially determines the bayesian network structure, in which the writing strategy is represented by a latent variable. By considering that each writing test is represented by its own (local) strategy and that there exists a global strategy which deals with each local strategy, we propose a Global Hierarchical Model. The EM learning algorithm can be used to obtain an unsupervised classifier that links handwriting and drawing features, local and global strategies. We give here preliminary results concerning a first phase of strategy discovering and labelling, and a second phase of observation of the pupil evolution in the discovered strategies.

Keywords

handwriting strategy, bayesian network, latent variables, clustering, EM algorithm.

1 Introduction

La maîtrise de l'écriture nécessite un apprentissage qui se superpose au développement du contrôle moteur de l'enfant [25]. Une hypothèse fondamentale dans la littérature affirme que "quelques caractéristiques de l'écriture peuvent être utilisées comme indicateur du niveau de développement".

Dans cet article, nous illustrons la notion de stratégie d'écriture à partir de caractéristiques mesurées pendant des épreuves précises et tentons de catégoriser des ensembles d'enfants ayant des caractéristiques proches en considérant qu'ils partagent la même stratégie d'écriture. Comme nous manipulons des variables latentes (i.e. les stratégies d'écriture, jamais mesurées), des données incertaines et/ou manquantes et des connaissances a priori fournies par des experts, nous avons choisi d'utiliser le formalisme des réseaux bayésiens pour modéliser notre problème. Les réseaux bayésiens sont des outils de représentation des connaissances et de raisonnement qui permettent de déterminer "graphiquement" les relations probabilistes entre les variables [22, 14, 15, 5]. Ces modèles probabilistes nous aident à estimer les relations entre les variables mesurées (caractéristiques des activités de contrôle moteur mesurées à partir de tests que les enfants effectuent sur une tablette graphique) et des stratégies d'écriture non mesurées.

Nous essaierons ensuite d'étiqueter les stratégies globales les plus fréquentes découvertes par notre modèle. Pour finir, nous étudierons l'évolution des enfants dans ces stratégies pendant les trois périodes d'acquisition effectuées.

2 Protocole expérimental

Prolongeant les travaux de [26, 23, 24, 25] et [3, 2, 7, 1], nous avons développé un logiciel d'acquisition en ligne de tests psychologiques concernant l'écriture et le dessin manuscrits. Ce logiciel extrait sept caractéristiques : vitesse moyenne (*vit*), écart-type de la vitesse (*ec*), nombre de traits (*trait*), durée des pauses (*paus*), accélération moyenne (*ac*), fluidité du mouvement (*fl*) et durée du tracé (*ex*). De précédentes expériences ont montré que, d'un point de vue cognitif, la fluidité est un bon indicateur du contrôle moteur, que le nombre de traits est lié à la spatialisation et que la durée de pause révèle des capacités de planification [30]. L'acquisition des données a eu lieu pendant trois périodes (T_1 , T_2 et T_3 , à des intervalles de 6 mois) avec les mêmes 102 enfants, sous les mêmes conditions expérimentales. Nous obtenons ainsi une base d'exemples de 306 mesures (102×3) de 42 caractéristiques (i.e. 7 variables pour 6 tests psychologiques).

Les tests psychologiques utilisés pendant ces expériences ont été proposés par [28] et [25] afin de tester différentes capacités (2 tests de dessins dérivés des travaux de Bender (cf. [29] et un exemple dans la figure 1) et 4 tests d'écriture : écrire son nom et son prénom, écrire un mot précis et isolé (cf. figure 1), et écrire une phrase fixée, avec puis sans le modèle). Les enfants sélectionnés, choisis du CP au

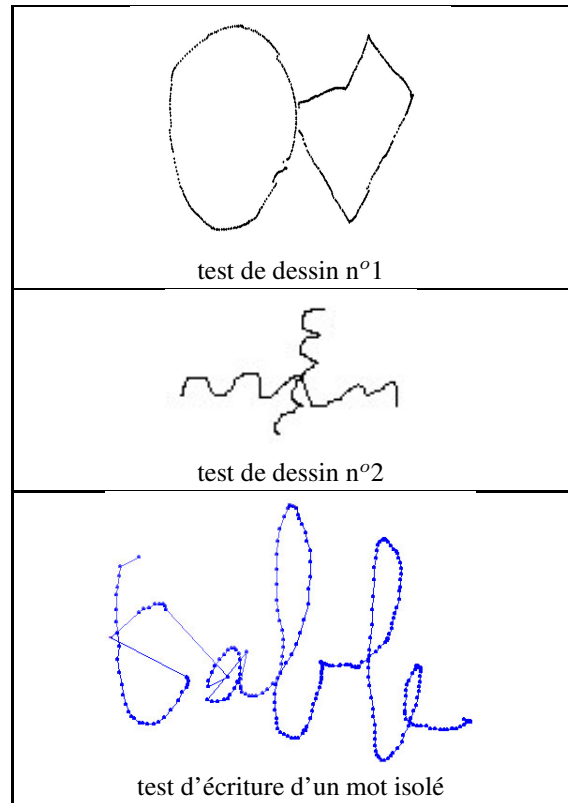


FIG. 1 – Exemple d'acquisition pour trois des six tests utilisés : deux tests de dessin (en haut) et un test d'écriture d'un mot isolé (en bas).

CM2 sont considérés comme des enfants "typiques" : ils n'ont jamais redoublé, n'ont aucun trouble du langage écrit ou oral et sont d'un âge normal pour leur niveau scolaire.

Le but de cette étude est d'essayer de découvrir des stratégies d'écriture communes à plusieurs de ces enfants en trouvant les liens entre les stratégies et les caractéristiques extraites par notre plateforme d'acquisition. Pour cela, nous proposons une approche basée sur le formalisme des réseaux bayésiens que nous allons maintenant décrire.

3 Création et apprentissage du réseau bayésien

L'utilisation d'un réseau bayésien pour modéliser un problème complexe passe par différentes phases. Nous commencerons par décrire l'architecture hiérarchique que nous utilisons puis détaillerons plus précisément le niveau local de notre modèle. L'architecture ainsi obtenue comporte un certain nombre de variables latentes (i.e. jamais mesurées), ce qui soulève deux questions auxquelles nous nous intéresserons : comment déterminer la taille des variables latentes et comment estimer les paramètres d'un modèle à partir de données incomplètes et/ou de variables latentes.

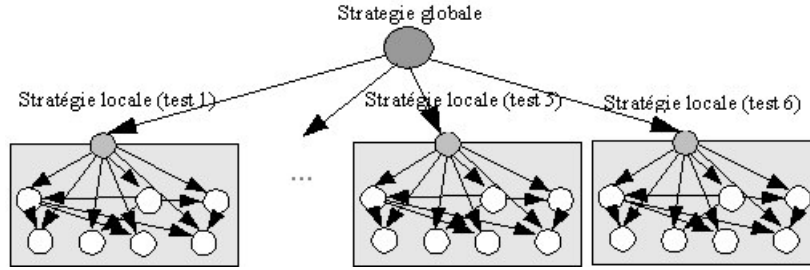


FIG. 2 – Modèle hiérarchique global : les stratégies locales correspondant aux 6 tests psychologiques dépendent toutes d’une stratégie globale (représentée par un rond gris).

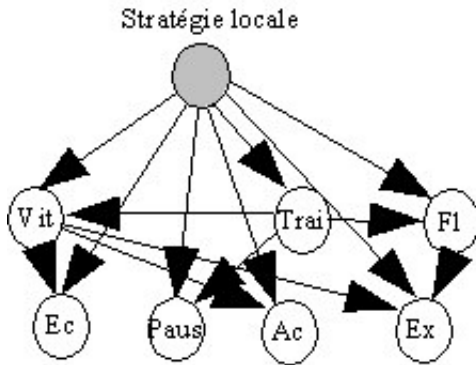


FIG. 3 – Modèle local : les relations entre les caractéristiques mesurées (représentées par des ronds blancs) sont données par des experts, mais ces variables dépendent toutes d’une stratégie locale (représentée par un rond gris).

3.1 Réseau bayésien hiérarchique

D’un point de vue cognitif, nous postulons que (1) chaque test psychologique utilisé dans nos expériences met en œuvre des capacités motrices spécifiques et que (2) toutes ces capacités motrices (locales) résultent d’un contrôle moteur (global). En nous inspirant des *hierarchical latent class models* proposés par [4] pour la visualisation de données ou [31] pour le clustering, nous modélisons notre problème avec le modèle hiérarchique de la figure 2 (Modèle Hierarchique Global) où les stratégies locales de chaque test sont des variables (discrètes) latentes qui sont indépendantes les unes des autres, mais qui dépendent toutes d’une stratégie globale (elle aussi latente). La valeur de cette stratégie globale correspond aux stratégies d’écriture que nous essayons de découvrir.

3.2 Structure locale

Considérons maintenant que chaque test psychologique est représenté par sa propre stratégie locale. Il faut déterminer la structure locale entre les variables mesurées (pour le test correspondant) et la stratégie locale. Plusieurs solutions s’offrent à nous. Une première solution est d’utiliser des algorithmes classiques d’apprentissage de structure dans les réseaux bayésiens [11]. Une autre solution est d’utiliser des structures "classiques" pour la classification

comme le réseau bayésien naïf. Une dernière approche est de déterminer la structure à partir de connaissances d’experts.

Comme nous possédons un nombre limité de données, la première approche est à écarter. Nous avons choisi d’utiliser une structure de réseau bayésien naïf augmenté souvent utilisée en reconnaissance des formes ([16, 27]). [13] a montré que ce genre de modèles est un bon compromis entre la simplicité du réseau bayésien naïf et les capacités du réseau bayésien classique à représenter des dépendances conditionnelles entre les variables. Dans une structure de réseau bayésien naïf augmenté, chaque variable mesurée possède la variable (cachée) stratégie comme parent mais peut aussi avoir comme parents d’autres variables mesurées. La structure augmentée que nous avons utilisée est décrite figure 3 et est issue de discussions avec les experts. Puisque la structure locale de la figure 3 est répétée pour chaque test psychologique (cf. figure 2), nous posons l’hypothèse que les relations entre les variables de cette structure locale ne dépendent pas des tests. Ce *partage* des probabilités conditionnelles nous permet de réduire le nombre de paramètres du réseau bayésien et par conséquent le nombre de données nécessaires pour l’apprentissage.

Notons aussi que les variables continues ont été discrétisées par une méthode issue de la théorie de l’information proposée par [6, 9] : les variables sont découpées en un certain nombre d’intervalles et les intervalles sont fusionnés lorsque la fusion ne fait pas perdre d’information.

3.3 Taille des variables latentes

La détermination de la taille d’une variable latente est un problème difficile [10]. Il est possible de l’estimer en utilisant des méthodes de validation croisée, en choisissant la taille de la variable qui maximisera un critère de score sur la structure obtenue, comme par exemple le critère BIC [33].

Malheureusement, l’utilisation de ces critères n’est pas forcément appropriée pour des modèles possédant des variables latentes. Des critères corrigés ont été proposés par [17] pour des *hierarchical latent class model* mais ils ne sont pas applicables dans le cas de notre architecture.

Récemment, [32] a proposé une adaptation de l’algorithme *Structural EM* [12, 21] pour les *Hierarchical Latent Class*

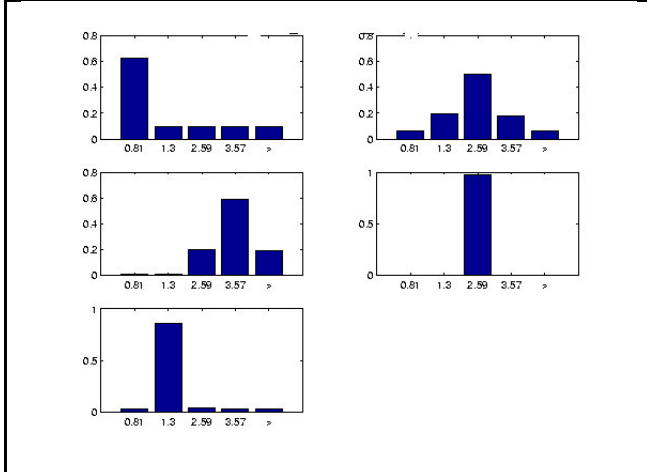


FIG. 4 – Distribution de probabilité de la vitesse moyenne (*vit*) pour le test d’écriture de mémoire, conditionnellement à la stratégie locale correspondante (*loc*=1 à 5). Nous pouvons remarquer que la première stratégie locale correspond à une vitesse faible, les stratégies 2 et 4 à des vitesses moyennes et la stratégie 3 à une vitesse élevée.

models. Cette adaptation essaie aussi d’optimiser la taille des variables latentes pendant l’apprentissage.

Nous avons alors préféré exploiter des travaux précédents et des résultats obtenus par une autre méthode de clustering (k-means) comme première approximation de la taille de nos variables cachées.

3.4 Apprentissage à partir de données incomplètes et/ou de variables latentes

Une fois la structure et la taille des variables latentes fixées, l’apprentissage des paramètres du réseau (i.e. les probabilités conditionnelles) à partir de données incomplètes et/ou de variables latentes est réalisé grâce à l’algorithme EM [8]. [18] a montré que cet algorithme s’adaptait naturellement aux réseaux bayésiens et autres modèles graphiques.

4 Résultats et interprétations

Après les phases de modélisation et d’apprentissage, nous allons maintenant tenter d’étiqueter les stratégies d’écriture découvertes à l’aide des distributions de probabilité de chaque variable mesurée conditionnellement aux stratégies d’écriture. Nous regrouperons ensuite les enfants selon ces stratégies et étudierons l’évolution temporelle des enfants dans ces stratégies pendant nos trois périodes d’acquisition.

4.1 Etiquetage des stratégies

Une fois le réseau bayésien appris, il est utilisé comme outil d’inférence. Il est ainsi possible de calculer et d’afficher la distribution de probabilité conditionnelle de n’importe quelle variable sachant que telle ou telle autre variable a été observée. Par exemple, la figure 4 nous donne la distribution de probabilité de la variable "vitesse moyenne" pour un des tests psychologiques (l’écriture de mémoire) selon

la valeur de la stratégie locale correspondante.

La distribution de probabilité a priori de la stratégie globale $P(G)$ nous dit que deux stratégies globales sont très courantes $G = 1$ (56%) et $G = 4$ (28%) alors que deux autres stratégies sont rares ($G = 2$ (7%) et $G = 4$ (9%)).

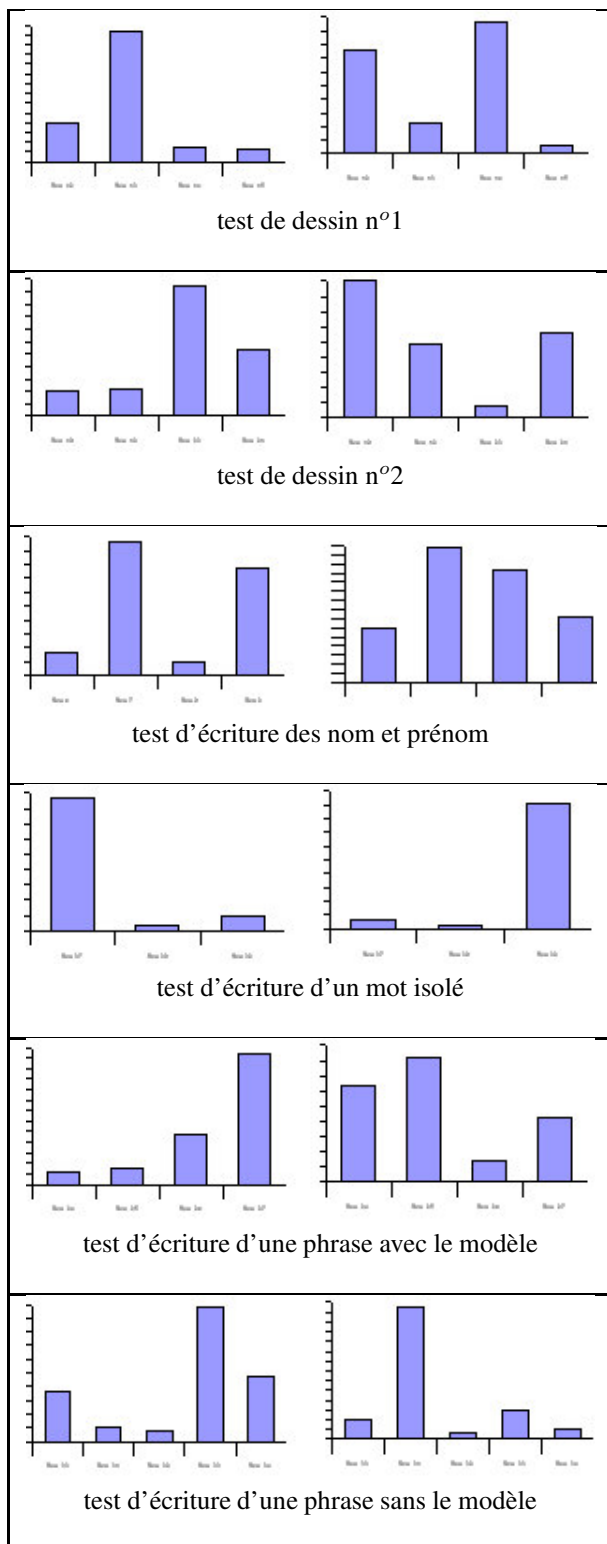
Nous avons donc commencé par essayer d’étiqueter les deux stratégies les plus probables à l’aide de psychologues. Cette étude nous a permis de découvrir que la stratégie globale $G = 4$ correspond à des *normo-scripteurs* (*NS*) et que la stratégie $G = 1$ représente des *normo-scripteurs avancés* (*NS+*). Par exemple, la stratégie *NS* correspond à des enfants ayant une fluidité moyenne pour la plupart des tests effectués, alors que les enfants appartenant à la stratégie *NS+* possèdent une meilleure fluidité. Cela rejoint les travaux de [30] qui affirment que l’anticipation peut être évaluée à partir de la fluidité du tracé.

Une première étude (cf. table 1) des stratégies non étiquetées ($G=2$ et $G=3$) nous indique qu’elles correspondent vraisemblablement à des stratégies locales opposées. Une étude plus précise au niveau des caractéristiques nous révèle de plus que $G=2$ correspond à un comportement au niveau de la spatialisation et de planification sensiblement supérieur que $G=3$ pour la plupart des tests psychologiques. Nous avons aussi noté que la stratégie $G=3$ rassemble des enfants ayant des stratégies locales similaires concernant trois des tests d’écriture : très bonne fluidité, faible nombre de traits et durée des pauses très courte. Afin d’aider à l’étiquetage de ces deux stratégies et pour confirmer l’étiquetage des deux stratégies découvertes, nous étudions maintenant l’évolution de stratégie des enfants au cours du temps.

4.2 Etude temporelle

La table 2 résume cette évolution : par exemple, 40 enfants sont regroupés dans la stratégie *NS* à T_1 , et 18 de ces 40 enfants restent dans cette stratégie en T_2 . Cette table nous montre aussi une propriété intéressante : la probabilité de transition entre *NS* et *NS+* et les probabilités de rester dans les stratégies *NS* ou *NS+* sont constantes au cours du temps. Par exemple, $P(NS \text{ à } T_2 | NS \text{ à } T_1) = 18/40 = 0.45$ et $P(NS \text{ à } T_3 | NS \text{ à } T_2) = 13/28 = 0.46$. La table 2 nous dévoile également que presque 46% des enfants conservent la stratégie *NS* entre chaque période d’acquisition, alors que 40% évoluent vers la stratégie *NS+*. Une fois dans la stratégie *NS+*, leur probabilité de ne pas changer de stratégie est d’environ 87%. Les enfants évoluent ainsi progressivement de la stratégie *NS* à la stratégie *NS+*.

La figure 5 représente la distribution selon le niveau scolaire (du CP au CM2) des enfants présents dans la stratégie *NS* (resp. *NS+*) à chaque période d’acquisition. Il est intéressant de remarquer que cette distribution selon le niveau scolaire est constante au cours du temps aussi bien dans la stratégie *NS* que dans la stratégie *NS+*. Nous pouvons aussi observer que les enfants classés dans la stratégie *NS* ont une faible probabilité d’être en CP, que cette



TAB. 1 – Distribution de la stratégie locale pour chacun des 6 tests psychologiques pour les stratégies globales non étiquetées ($G=2$ à gauche et $G=3$ à droite). Ces deux stratégies correspondent vraisemblablement à des stratégies locales opposées.

		Stratégie globale (T_2)			$Tot(T_1)$
		NS	$G = 2$	$G = 3$	
Stratégie globale (T_1)	NS	18	4	2	16
	$G = 2$	2			1
	$G = 3$	5	3	3	4
	$NS+$	3	3		38
$Tot(T_2)$		28	10	5	59

		Stratégie globale (T_3)			$Tot(T_2)$
		NS	$G = 2$	$G = 3$	
Stratégie globale (T_2)	NS	13	1	2	12
	$G = 2$	4	3		3
	$G = 3$				5
	$NS+$	3	1	3	52
$Tot(T_3)$		20	5	5	72

TAB. 2 – Evolution de la stratégie de nos 102 enfants entre T_1 et T_2 (haut) et entre T_2 et T_3 (bas). Parmi les 40 enfants classés dans la stratégie NS en T_1 , 18 restent dans la même stratégie pendant que 16 passent en $NS+$. Au même moment, 38 des 44 enfants classés dans la stratégie $NS+$ restent dans la même stratégie. Notons que l'évolution est proportionnellement la même entre T_2 et T_3 . $G=2$ et $G=3$ semblent être des stratégies intermédiaires mais ne sont pas encore étiquetées et devront faire l'objet d'une étude plus précise.

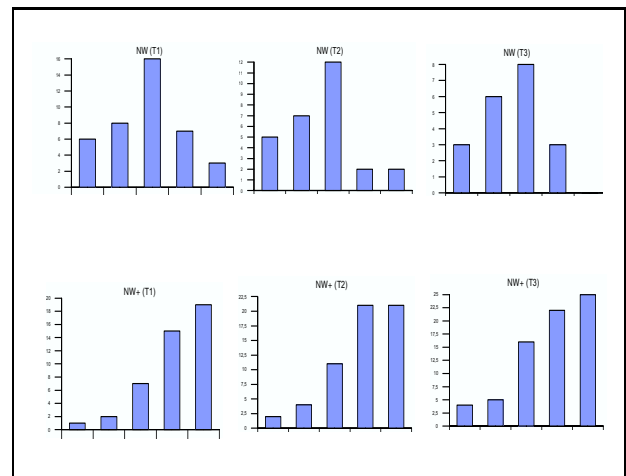


FIG. 5 – Distribution de probabilité des 5 niveaux scolaires des enfants (CP à CM2) classés dans la stratégie NS (ligne du haut) et dans la stratégie $NS+$ (ligne du bas) pour chacune des périodes d'acquisition T_1 , T_2 et T_3 (en colonnes).

probabilité augmente jusqu'au CE2 puis décroît jusqu'au CM2. Aussi logiquement, pour les enfants de la stratégie $NS+$, la probabilité d'être au CP est quasiment nulle, puis augmente très fortement en CM1 et CM2.

La table 2 nous permet également de déduire que ces stratégies sont des stratégies intermédiaires. Seuls quelques enfants sont classés dans ces stratégies, et ils y restent pendant peu de temps, rejoignant rapidement une des stratégies éti-

quetées (NS ou $NS+$). Les expériences actuelles ne nous permettent pas d'étiqueter clairement ces deux stratégies. Une étude plus précise devra être menée afin d'en savoir plus.

5 Conclusion

Nous avons décrit dans cet article une application des réseaux bayésiens à la modélisation et la découverte de stratégies d'écriture chez les enfants en scolarité primaire, à partir de caractéristiques extraites automatiquement pendant des tests psychologiques. Notre modèle combine plusieurs modélisations typiques des réseaux bayésiens utilisés en classification : structure naïve augmentée pour la partie locale de notre architecture, et arborescence de variables latentes pour la partie globale.

Ce modèle hiérarchique, appris à partir de données réelles, nous a ensuite permis de découvrir deux stratégies globales qui correspondent à des comportements représentatifs de normo-scripteurs (NS) et normo-scripteurs avancés ($NS+$).

L'étude suivante a concerné l'évolution des enfants dans ces stratégies pendant trois périodes d'acquisition successives. Cette étude nous a permis de vérifier que nos hypothèses concernant NS et $NS+$ étaient consistantes : la distribution selon le niveau scolaire correspondant à chacune de ces stratégies est constante au cours du temps, et les probabilités de transition (ou de rester dans le même état) ne dépendent pas non plus du temps.

Les deux autres stratégies globales $G=2$ et $G=3$ concernent un nombre réduit d'enfants et apparaissent comme des stratégies intermédiaires. Une étude plus précise devra être menée afin d'en savoir plus et d'étiqueter clairement ces deux stratégies, en examinant en détail avec des experts les stratégies locales correspondant à chaque test ainsi que les résultats de ces tests pour les enfants classés dans ces stratégies non étiquetées.

Notre modèle utilise un certain nombre d'avantages des réseaux bayésiens : apprentissage de probabilités conditionnelles à partir de données incomplètes et incertaines, prise en compte d'une expertise pour la détermination partielle de la structure du réseau, utilisation de l'inférence pour afficher les distributions de probabilités conditionnelles de n'importe quelle variable dans le réseau afin d'aider les experts à qualifier notre modèle.

En application directe, pour des résultats de tests donnés, il est possible d'estimer quelle est la stratégie de l'enfant, ou éventuellement si l'enfant ne suit aucune des stratégies connues. Il est possible aussi de raisonner "à l'envers", en se fixant une stratégie d'écriture, et en étudiant la capacité de chaque test psychologique à détecter la stratégie correspondante.

Concernant les perspectives, la matrice de transition de la table 2 semblait sensiblement constante au cours du temps. Une extension intéressante serait de considérer maintenant notre modèle hiérarchique global (figure 2) comme la partie statique du réseau bayésien dynamique décrit figure 6

(dont le principe est similaire aux modèles de Markov cachés). Des résultats préliminaires tendent à montrer que ce type de réseau dynamique pourrait modéliser l'évolution des enfants dans les stratégies et pourrait être utilisé pour inférer par exemple les résultats à tel ou tel test à l'instant $t + 1$ sachant les résultats à l'instant t .

Remerciement

Toutes nos expériences ont été réalisées avec BNT, toolbox réseaux bayésiens pour Matlab proposée par [20] et dont vous pouvez trouver plus d'informations sur le site français [19].

Références

- [1] M. Amara. *Segmentation de Tracés Manuscrits Application à l'Extraction de Primitives*. PhD thesis, Université de Rouen, 1998.
- [2] M. Amara, P. Courtellemont, D. de Brucq, and R. Devinoy. An analysis software tool for handwriting : Writing and drawing applications. In *Proceedings of the 8th biennial conference of the International Graphonomics Society, IGS'97*, pages 107–108, Gènes, Italie, 1997.
- [3] M. Amara, P. Wallon, D. de Brucq, C. Mesmin, and R. Baldy. Segmentation de tracés manuscrits par des segments de droite : application à l'extraction de segments de droite sur la figure de rey. approche préliminaire. In *10ème Congrès en Reconnaissance de formes et Intelligence Artificielle, RFIA' 96, Vol. 1*, pages 456–463, Rennes, France, 1996.
- [4] C.M. Bishop and M.E. Tipping. A hierarchical latent variable model for data visualisation. *IEEE T-PAMI*, 3(20) :281–293, 1998.
- [5] C. Borgelt and R. Kruse. *Graphical Models - Methods for Data Analysis and Mining*. John Wiley & Sons, Chichester, United Kingdom, 2002.
- [6] O. Colot, C. Olivier, P. Courtellemont, and A. El-Matouat. Information criteria and abrupt changes in probability laws. *Signal Processing VII : Theory and Applications*, pages 1855–1858, 1994.
- [7] D. de Brucq, M. Amara, and V. Ruiz. On-line handwritten segmentation in linear drawings. *Pattern Recognition Letters*, 18 :193–202, 1997.
- [8] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B* 39 :1–38, 1977.
- [9] F. El Matouat, O. Colot, P. Vannoorenberghe, and J. Labiche. From continuous to discrete variables for bayesian network classifiers. In *Conference on Systems, Man and Cybernetics, IEEE-SMC*, Nashville, USA, 2000.

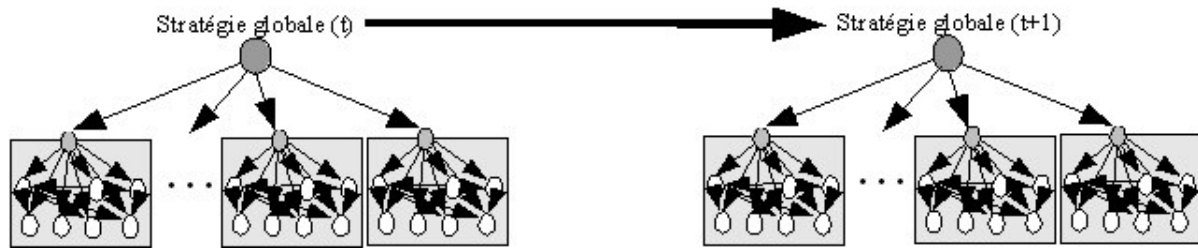


FIG. 6 – Amélioration du modèle hiérarchique global pour modéliser l'évolution temporelle entre les stratégies globales.

- [10] G. Elidan and N. Friedman. Learning the dimensionality of hidden variables. In *Uncertainty in Artificial Intelligence : Proceedings of the Seventeenth Conference (UAI-2001)*, pages 144–151, San Francisco, CA, 2001. Morgan Kaufmann Publishers.
- [11] O. Francois and P. Leray. Etude comparative d'algorithmes d'apprentissage de structure dans les réseaux bayésiens. In *Proceedings of RJCIA 2003, plateforme AFIA 2003*, Laval, France, 2003.
- [12] N. Friedman. The Bayesian structural EM algorithm. In Gregory F. Cooper and Serafín Moral, editors, *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98)*, pages 129–138, San Francisco, July 24–26 1998. Morgan Kaufmann.
- [13] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29 :131 – 163, 1997.
- [14] F.V. Jensen. *Introduction to Bayesian Networks*. Springer Verlag, 1996.
- [15] M.I. Jordan. *Learning in Graphical Models*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
- [16] E. Keogh and M. Pazzani. Learning augmented bayesian classifiers : A comparison of distribution-based and classification-based approaches. In *Proceedings of the Seventh International Workshop on Artificial Intelligence and Statistics*, pages 225–230, 1999.
- [17] T. Kocka and N. L. Zhang. Dimension correction for hierarchical latent class models. In Adnan Darwiche and Nir Friedman, editors, *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI-02)*, pages 267–274, S.F., Cal., 2002. Morgan Kaufmann Publishers.
- [18] S. L. Lauritzen. The EM algorithm for graphical association models with missing data. *Computational Statistics and Data Analysis*, 1 :191–201, 1995.
- [19] P. Leray, S. Guilmineau, G. Noizet, and O. Francois. French BNT site, 2003. <http://bnt.insa-rouen.fr/>.
- [20] K. Murphy. The BayesNet Toolbox for Matlab, *Computing Science and Statistics : Proceedings of Interface*, 33, 2001. <http://www.ai.mit.edu/~murphyk>.
- [21] J.M. Peña, J.A. Lozano, and P. Larrañaga. An improved bayesian structural em algorithm for learning bayesian networks for clustering. *Pattern Recognition Letters*, 21 :779–786, 2001.
- [22] J. Pearl. *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. Morgan Kaufmann, 1988.
- [23] C. Rémi. *Contribution au choix de primitives en graphomotricité*. PhD thesis, Université de Rouen, 1999.
- [24] C. Rémi, M. Amara, P. Courtellement, D. de Brucq, and R. Devinoy. Un logiciel d'analyse automatique de la structuration de figures géométriques acquises en ligne. In *VI'99, 12th Canadian conf. devoted to computer Vision, signal and Image processing and pattern recognition*, pages 233–240, Québec, Canada, 1999.
- [25] C. Rémi, C. Frélicot, and P. Courtellement. Automatic analysis of the structuring of children's drawing and writing. *Pattern Recognition*, 35(5) :1059–1069, 2002.
- [26] C. Rémi, M. Amara, P. Courtellement, and D. de Brucq. An experimental program for the automatic study of handwriting learning at elementary school. In *Proceedings of the 8th biennial conference of the International Graphonomics Society, IGS'97*, pages 75–76, Gènes, Italie, 1997.
- [27] J.P. Sacha, L. Goodenday, and K.J. Cios. Bayesian learning for cardiac spect image interpretation. *Artificial Intelligence in Medicine*, 26 :109–143, 2002.
- [28] P. Westzaan and J. Kosterman. Diagnosis and evaluation of fine motor behaviour in children. In *Proceedings of ICOD'93*, 1993.
- [29] R. Zazzo. Test moteur de structuration visuelle de bender. In *Manuel pour l'examen psychologique de l'enfant, vol.1*. Delachaux et Niestle ; Actualités pédagogiques et psychologiques, 1992.
- [30] P. Zesiger. *Approches cognitive, neuropsychologique et développementale*. Presses Universitaire de France, 1995.
- [31] N. L. Zhang. Hierarchical latent class models for cluster analysis. In *Proceedings of AAAI'02*, 2002.

- [32] N. L. Zhang. Structural em for hierarchical latent class model. Technical report, HKUST-CS03-06, 2003.
- [33] N. L. Zhang, T.D. Nielsen, and F.V. Jensen. Latent variable discovery in classification models. *to appear in Artificial Intelligence in Medicine*, 2003.