

Républiqu Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Ferhat Abbas Sétif 1
Faculté des Sciences



MEMOIRE DE MASTER

Domaine: Math et Informatique
Département: Informatique
Specialité: Informatique fondamentale et intelligence artificielle

Thème

Classification hybride (Hard et Soft) d'items éducatifs similaires

Présenté par :
Abdou Abarchi Aboubacar

Encadré par :
M^{me} **Harbouche Khadidja**

Devant le jury composé de :

Président : *M^r* **Abdellatif HABES**
Examinatrice : *M^{me}* **Houda HAFI**

Juillet 2021

**M
A
S
T
E
R**

Table des matières

Table des figures	vii
Liste des tableaux	viii
Liste des Acronymes	viii

I État de l'art	1
1 Educational data minig	2
1.1 Introduction	3
1.2 Data mining	3
1.2.1 Définition	3
1.2.2 Les méthodes du data mining	4
1.3 Intelligent Tutorial Systems (ITS)	4
1.3.1 Définition	4
1.3.2 ITS architecture	5
1.4 Educational Data mining	6
1.4.1 Définition	6
1.4.2 Processus d'application du DM en éducation	7
1.4.2.1 Les méthodes d'analyse et d'exploration appliquer dans EDM	8
1.4.2.1.1 L'induction de modèle supervisée	9
1.4.2.1.2 L'induction de modèle non supervisée	9
1.4.2.1.3 L'estimation des paramètres	10
1.4.2.1.4 L'exploration de relations (Relationship mining)	10
1.4.2.1.5 La distillation des données pour le jugement humain (Distillation of data for human judgment)	10
1.4.2.1.6 Discovery with models	11
1.5 Conclusion	11
2 Modèle de l'apprenant et découverte des prérequis	12
2.1 Introduction	13
2.2 Learner Model	13
2.2.1 Définition	13
2.2.2 Caractéristiques des méthodes utilisées pour la modélisation de l'apprenant	13
2.2.3 Composantes du modèle de l'apprenant	16
2.2.3.1 Modèle cognitif	17
2.2.3.2 Modèle d'inférence	17
2.2.3.3 Modèle émotionnel	17

2.2.4	Utilitaire du modèle de l'apprenant	18
2.2.5	Types de modèles d'apprentissage	18
2.2.6	Contenu du modèle de l'apprenant	18
2.3	Modèle de compétence	19
2.3.1	Granularité	19
2.3.2	Relations pré-requises	20
2.4	Découverte des prérequis	22
2.5	Conclusion	23
3	L'approche items-to-skills mapping	24
3.1	Introduction	25
3.2	Conclusion	25
4	Fiabilité et l'accord pour les données d'observations	26
4.1	Introduction	27
4.2	Conclusion	27
5	Inférence bayésienne et la théorie de la réponse aux items	28
5.1	Introduction	29
5.2	Conclusion	29
6	Clustering	30
6.1	Introduction	31
6.2	Conclusion	31
II	Contributions	32
7	Contribution and Results	33
7.1	Introduction	34
7.2	Conclusion	34
	Conclusion Générale	34
	Bibliographie	36

Table des figures

1.1	Les étapes de l'extraction de connaissances à partir de données	4
1.2	ITS architecture	5
1.3	Principaux domaines liés à l'exploration de données éducatives	7
1.4	Data mining application process applied in education	8
1.5	EDM techniques	9
2.1	Composantes du modèle de l'apprenant	16
2.2	Deux alternatives pour modéliser les relations d'agrégation.	20
2.3	Agrégation de modélisation de réseau bayésien et relations préalables simultanément	21
2.4	Modèles d'équations structurelles	22

Liste des tableaux

2.1	Caractéristiques des méthodes utilisées pour la modélisation des apprenants partie[1]	14
2.2	Caractéristiques des méthodes utilisées pour la modélisation des apprenants partie[2]	15

Liste des Acronymes

EDM	Educational Data Mining
ITS	Intelligent Tutorial Systems
IALT	Intelligent Computer Assisted Instruction
CAI	Computer Assisted Instruction
BKT	Bayesian Knowledge Tracing

Première partie

État de l'art

CHAPITRE 1

Educational data minig

Sommaire

1.1	Introduction	3
1.2	Data mining	3
1.3	Intelligent Tutorial Systems (ITS)	4
1.4	Educational Data mining	6
1.5	Conclusion	11

1.1 Introduction

Du fait que les systèmes informatisés conservent des journaux détaillés des interactions utilisateur-système, ces journaux détaillés qui sont généralement dans une grande base de données ouvrent de nouvelles opportunités pour étudier ces données récolter.

L'EDM emprunte et étend des domaines connexes tels que l'apprentissage automatique (l'étude des programmes informatiques qui apprennent et s'améliorent avec des données empiriques), l'exploration de texte (approches pour trouver des modèles dans texte en langue) et statistiques. D'autres influences importantes sont la psychométrie (l'étude des instruments psychologiques pour mesurer les compétences et les traits humains) et l'analyse des journaux Web (approches pour identifier les profils d'utilisateurs et la navigation modèles d'utilisateurs du site Web).

Dans ce chapitre, nous présentons les techniques d'exploration de données, et un ensemble de technique d'analyse pour une variété de problèmes dans la recherche en éducation et développement technologique (Romero et al. 2010 ; International Working Group on Educational Data Mining, n. d.) [1].

1.2 Data mining

1.2.1 Définition

"L'exploration de données est un processus itératif et interactif dont l'objectif est la découverte de modèles de données valides, nouveaux, utiles et compréhensibles dans de grandes bases de données." (Talbi, 2015)

L'exploration de données, également appelée découverte de connaissances dans une base de données, fait référence à l'extraction de connaissances à partir de grandes quantités de données. Elle est utilisée pour extraire de grands volumes de données pour découvrir des modèles cachés et des relations utiles dans la prise de décision. Bien que l'exploration de données et la découverte de connaissances de base de données soient souvent traitées comme des synonymes, elles font en fait partie du processus de découverte des connaissances. La séquence des étapes identifiées lors de l'extraction des connaissances à partir des données est illustrée à la figure 1.1 [2].

L'exploration de données est un domaine interdisciplinaire utilisant à la fois des techniques

d'apprentissage automatique, de reconnaissance de formes et l'utilisation de statistiques, de bases de données et de visualisation pour déterminer les façons d'utiliser l'extraction d'informations à partir de très grandes bases de données (Cabena et al., 1998).

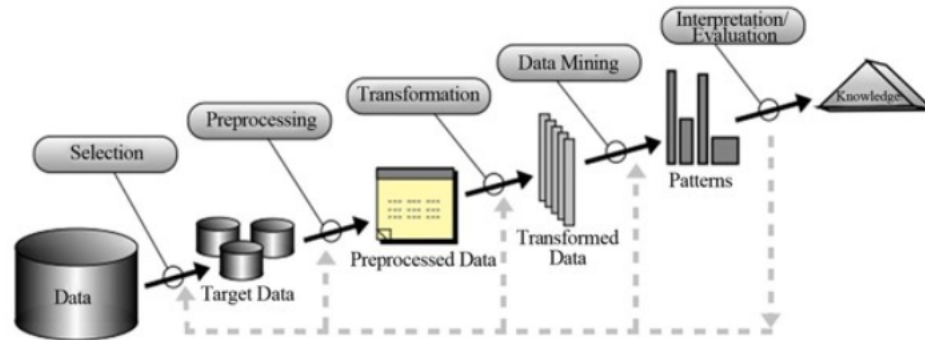


FIGURE 1.1 : Les étapes de l'extraction de connaissances à partir de données

1.2.2 Les méthodes du data mining

Le clustering et la classification sont deux tâches fondamentales dans l'exploration de données. La classification est principalement utilisée comme méthode d'apprentissage supervisé, le clustering pour l'apprentissage non supervisé (certains modèles de clustering sont pour les deux). Le but du clustering est descriptif, celui de la classification est prédictif (Veyssieres et Plant, 1998).

« Comprendre notre monde nécessite de conceptualiser les similitudes et les différences entre les entités qui le composent » (Tyron et Bailey, 1970) [3]. De ce fait, certains coefficients de corrélations sont utilisés pour calculer la similarité entre deux éléments ou plusieurs et, des métriques et les critères de liaison permettent qu'en ta eux de déterminer le lien entre plusieurs entités afin de les groupées dans une même classe (c'est le cas du clustering).

Certaines méthodes comme : rule of association, preaching, neural networks, decision trees sont utilisé aussi pour explorer et analyser des jeux de données.

1.3 Intelligent Tutorial Systems (ITS)

1.3.1 Définition

Les systèmes de tutorat intelligents (ITS) sont des environnements d'apprentissage informatisés issus de l'IALT (Intelligent Computer Assisted Instruction) qui visent à personnaliser la formation. En effet, ils ont été développés pour répondre aux limites du CAI (Computer

Assisted Instruction) en utilisant l'intelligence artificielle pour mettre en œuvre des systèmes plus flexibles et interactifs qui s'adaptent « aux besoins spécifiques de l'apprenant en évaluant et en diagnostiquant ses problèmes afin de fournir les assistances » (Buche C., 2005). L'objectif est de mimer le comportement d'un tuteur humain en sa qualité de pédagogue expert et d'expert dans le domaine. Ainsi, tout comme un tuteur, un logiciel de ce type a le potentiel d'amener l'apprenant à accomplir une tâche et à fournir un retour sur leurs actions. ITS répond ainsi à la nécessité de placer l'apprenant au centre du processus d'apprentissage [4].

Les ITS sont essentiellement un environnement de résolution de problèmes ou d'exercice. Ils favorisent l'apprentissage dans un domaine spécifique en guidant et en aidant l'apprenant. Parfois, ils exposent d'abord le contenu de la matière à l'apprenant ; parfois ils présentent directement les exercices qui permettront à l'apprenant d'assimiler les connaissances [5].

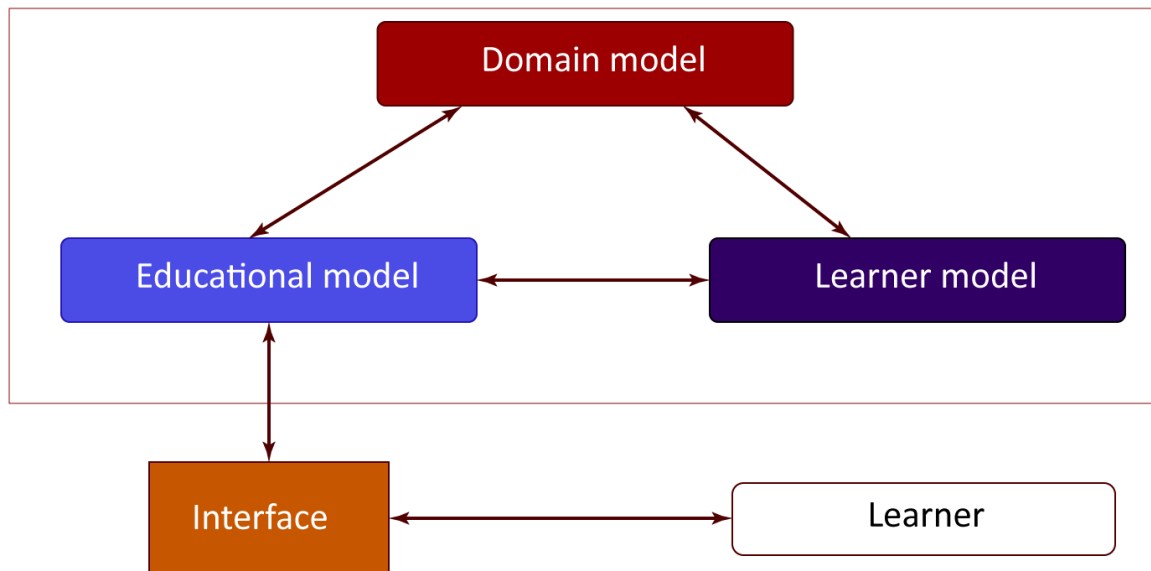


FIGURE 1.2 : ITS architecture

1.3.2 ITS architecture

L'architecture conceptuelle d'un ITS est illustrée à la figure 1.2 et comprend : [5]

- **Un modèle de l'interface** : qui représente la couche de communication entre l'apprenant et le système, c'est-à-dire les interactions entre l'apprenant et le système. Cette interface peut faire varier le type d'environnement d'apprentissage et l'objectif étant de privilégier une approche de conception qui ne gêne pas l'apprentissage. La qualité de l'interaction peut influencer les résultats d'apprentissage. Par conséquent, les principaux problèmes avec les interfaces d'apprentissage sont les problèmes d'inter-

action personne / machine. Pour surmonter ces problèmes, il faut prêter attention à la convivialité pour que la charge mentale associée à l'interface soit négligeable, ainsi qu'à l'utilité en facilitant l'accès, en permettant l'accès au domaine d'apprentissage et en soutenant la métacognition de l'apprenant. Les tendances actuelles s'orientent vers des dialogues tutoriels en langage naturel, l'intégration de la dimension affective dans l'interaction et les interfaces tangibles.

- **Modèle du domaine** : également appelé modèle expert, il représente l'expertise de l'enseignant dans le domaine, c'est-à-dire ce qui doit être enseigné. Elle peut être modélisée de trois manières : une approche « boîte noire » : appliquer toute méthode de raisonnement sur le domaine inexplicé, sans aucune transparence pour l'utilisateur ; un système expert peut expliquer son raisonnement et un modèle cognitif : simuler la manière dont les humains utilisent les connaissances.
- **Un modèle d'apprentissage** : qui peut personnaliser l'apprentissage en tenant compte des spécificités de l'apprenant.
- **Un modèle éducatif** : qui sait enseigner.

1.4 Educational Data mining

1.4.1 Définition

La communauté d'exploration de données appliquée dans l'éducation a défini ce terme comme suit : « L'exploration de données appliquée dans l'éducation est une discipline qui concerne le développement de méthodes permettant d'explorer les types uniques de données provenant des milieux éducatifs. Ces méthodes sont utilisées pour mieux comprendre le comportement des apprenants et environnement de leur apprentissage. » [6]

L'exploration de données éducatives (EDM) vise à développer, rechercher et appliquer des méthodes informatisées pour détecter des modèles dans de grandes collections de données éducatives qui seraient autrement difficiles ou impossibles à analyser en raison de l'énorme volume de données dans lesquelles elles existent [3].

Ce domaine est une forme d'intersection des trois domaines principaux tels que : l'informatique, l'éducation et les statistiques illustrés dans la figure 1.3.

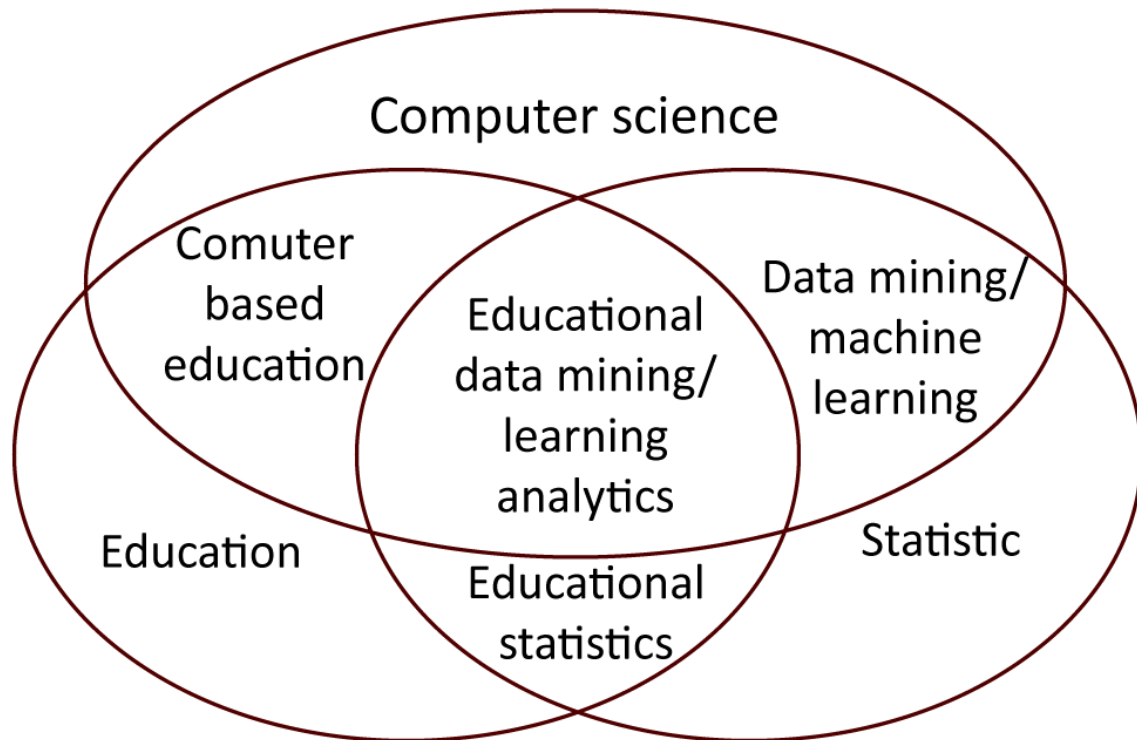


FIGURE 1.3 : Principaux domaines liés à l’exploration de données éducatives

Les techniques d’exploration de données (EDM) sont devenues plus importantes dans la recherche et le développement en raison du développement rapide de la technologie, de la croissance rapide des connaissances humaines et de l’augmentation du nombre de personnes détenant des systèmes d’enseignement informatisé.

1.4.2 Processus d’application du DM en éducation

Le processus d’application de l’exploration de données appliqué dans l’éducation peut être considéré comme un cycle itératif de formulation, de test et de raffinement d’hypothèses. Dans ce processus, l’objectif n’est pas seulement de transformer les données en connaissances, mais aussi de filtrer les connaissances extraites pour savoir comment les modifier.

L’environnement éducatif pour améliorer l’apprentissage des apprenants est présenté dans la figure 1.4. Des études ont montré que l’application de l’EDM est similaire au processus Knowledge from Data (KDD). Ce processus commence par la collecte de données à utiliser à partir de l’environnement éducatif.

Les données brutes obtenues nécessitent un nettoyage et un prétraitement tels que : fusion de données hétérogènes, traitement des données manquantes, conversion de données d’une

source de données à une autre, et données, etc. Cette phase nécessite souvent l'utilisation de certaines données techniques minières.

Le résultat de ce dernier est un modèle capable de structurer les données stockées. Enfin, la dernière étape est l'interprétation et l'évaluation des résultats obtenus [7].

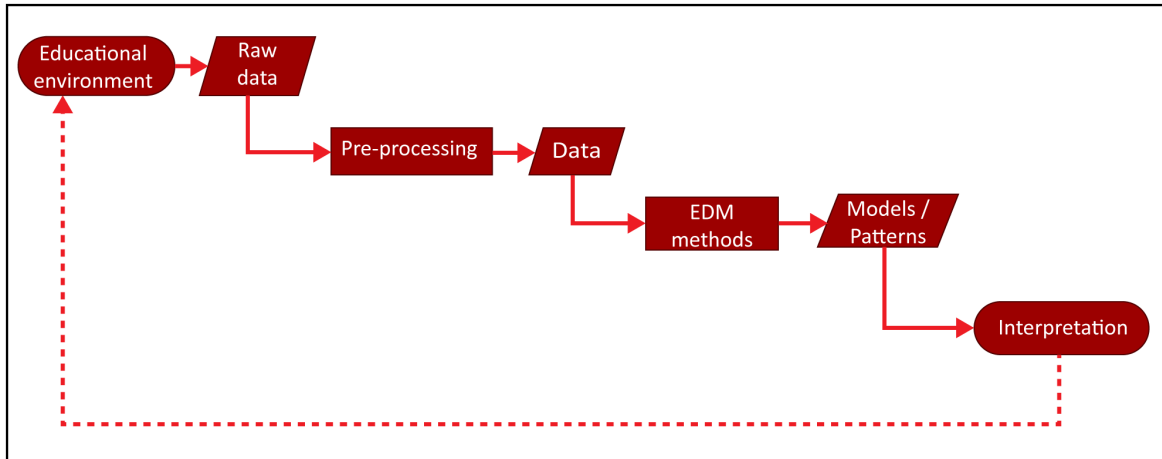


FIGURE 1.4 : Data mining application process applied in education

1.4.2.1 Les méthodes d'analyse et d'exploration appliquer dans EDM

Certaines méthodes du data mining ont été largement appliquées sur les données issues des systèmes éducatif afin d'obtenir des connaissances cachées et des modèles. Les méthodes du data mining ont été appliquées dans plusieurs domaines de recherche tels que : le e-learning, le système de tuteur intelligent, text mining, réseaux sociaux, web mining, etc. Les méthodes d'analyse et d'exploration de données éducatives sont tirées de diverses sources littéraires, notamment l'apprentissage automatique, la psychométrie et d'autres domaines de la modélisation informatique, des statistiques et de la visualisation de l'information.

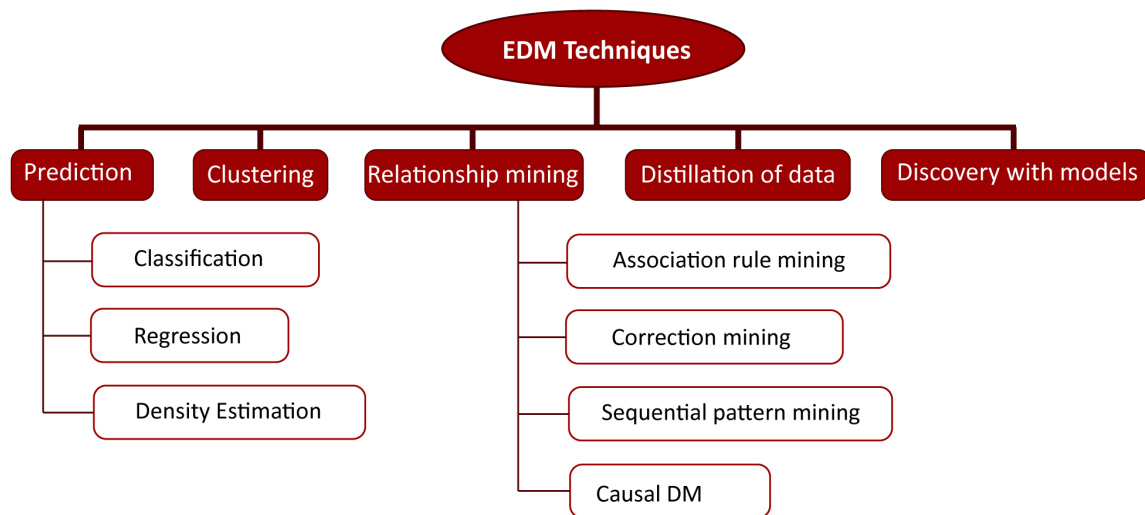


FIGURE 1.5 : EDM techniques

1.4.2.1.1 L'induction de modèle supervisée L'induction de modèle supervisée comprend des techniques d'apprentissage automatique qui déduisent des modèles de prédiction à partir de instances d'entraînement pour lesquelles les valeurs d'un attribut cible sont connues. Les modèles de prédiction acceptent instances en entrée (généralement décrites comme un vecteur d'attribut) et en sortie une prédiction pour la cible attribut. Les modèles qui prédisent des valeurs cibles catégorielles sont appelés modèles de classification; des modèles qui prédire les valeurs cibles continues sont appelés modèles de régression. Les modèles de prédiction peuvent être basés sur différentes représentations, par exemple, les arbres de décision, les machines à vecteurs de support (les deux classifications) et modèle de régression linéaire (régression) [1].

1.4.2.1.2 L'induction de modèle non supervisée L'induction de modèle non supervisée comprend des techniques d'apprentissage automatique qui déduisent des modèles à partir d'instances d'apprentissage pour lesquelles les valeurs d'un attribut cible ne sont pas connues. Les méthodes non supervisées utilisent une approche ascendante, c'est-à-dire que les modèles et les structures sont recherchés dans l'espace d'entrée sans catégories cibles explicitement définies ni exemples étiquetés. Une approche largement utilisée est le clustering, qui est utilisé pour identifier des groupes d'instances dans un ensemble d'apprentissage qui sont « similaires » à certains égards. En règle générale, une sorte de mesure de distance (par exemple, la distance euclidienne) est utilisée pour décider de la similitude des instances. Une fois un ensemble de clusters a été déterminé, de nouvelles instances peuvent être classées en déterminant le cluster le plus proche. Un algorithme de clustering bien connu est le clustering k-means [1].

1.4.2.1.3 L'estimation des paramètres L'estimation des paramètres comprend des techniques statistiques pour déduire des paramètres de modèles probabilistes à partir d'un jeu de données. Ces modèles peuvent être utilisés pour prédire la probabilité d'événements d'intérêt. L'approche est basée sur l'hypothèse que le modèle a une forme paramétrique donnée (par exemple, une distribution gaussienne avec les paramètres moyenne et variance). Un exemple d'application en EDM est l'estimation des paramètres Bayesian Knowledge Tracing (BKT). Le BKT est utilisé pour déterminer la probabilité qu'un élève maîtrise une compétence sur la base de l'historique des performances passées. Un modèle BKT peut être compris comme un réseau bayésien dynamique avec quatre paramètres (priorité, estimation, glissement et taux d'apprentissage). Ces paramètres peuvent être déterminés, par exemple, avec l'algorithme Espérance-Maximisation. Aussi l'utilisation de la théorie des réponses aux items (Item Response Theory IRT) permet de capter plus de nuances dans le comportement humain et d'estimer la capacité de l'apprenant à réussir un item et aussi d'estimer la difficulté, la discrimination des items [1].

1.4.2.1.4 L'exploration de relations (Relationship mining) L'exploration de relations concerne l'identification des relations entre les variables, des relations qui peuvent être de nature associative, corrélationnelle, séquentielle ou causale. Par exemple, une approche courante de l'exploration de règles d'association consiste à apprendre les règles SI-ALORS qui dépassent un seuil minimum de « support » et de « confiance ». La prise en charge indique la fréquence relative des transactions qui correspondent à la fois aux parties IF et ALORS de la règle. La confiance dénote la fréquence relative des transactions qui correspondent à la partie THEN de la règle dans l'ensemble de transactions qui correspondent à la partie IF. Apriori est, par exemple, un algorithme de règle d'association classique. Un exemple d'application en EDM est l'identification d'erreurs qui se produisent fréquemment ensemble (par exemple, les étudiants qui ont commis les erreurs A et B ont également souvent commis l'erreur C) [1].

1.4.2.1.5 La distillation des données pour le jugement humain (Distillation of data for human judgment) La distillation des données pour le jugement humain vise à représenter les données de manière intelligible à l'aide de statistiques, méthodes de visualisation et interfaces d'information interactives. Par exemple, les performances moyennes les scores peuvent être calculés pour chaque élève et présentés à un enseignant par ordre croissant dans un graphique à barres. Un autre exemple est les courbes d'apprentissage, qui tracent les performances d'un élève (par exemple, le temps de réponse) par rapport au nombre d'occasions de pratiquer une compétence. Une courbe d'apprentissage idéale montre que la

performance s'améliore en douceur et de façon monotone, suivant approximativement une loi de puissance ou une fonction exponentielle. D'un autre côté, les courbes d'apprentissage avec des pointes indiquent qu'une autre compétence pourrait interférer avec la compétence réellement modélisée, c'est-à-dire que le modèle de compétence pourrait être amélioré [1].

1.4.2.1.6 Discovery with models Discovery with models (découverte avec des modèles) comprend des approches qui amorcent des modèles déjà existants pour faire des découvertes plutôt que de calculer de nouveaux modèles à partir de zéro. Par exemple, un modèle de prédiction pourrait être appliqué à un ensemble de données pour prédire les valeurs d'une catégorie cible d'intérêt. Les prédictions elles-mêmes pourraient être utilisés à nouveau comme données dans d'autres analyses, par exemple, ils pourraient être corrélés avec une catégorie cible d'un autre modèle de prédiction. Un autre exemple consiste à scruter les différentes composantes d'un modèle de prédiction pour en savoir plus sur les facteurs qui influencent la prédiction [1].

1.5 Conclusion

Dans ce chapitre, nous avons passé en revue le domaine de l'exploration de données éducatives, en présentant brièvement ses applications et ses techniques. L'application des méthodes d'exploration de données dans le secteur de l'éducation est un phénomène intéressant. Les techniques d'exploration de données dans les organisations éducatives nous aident à en apprendre davantage sur les performances des apprenants, le comportement des apprenants, la conception des programmes et à motiver les apprenants sur divers paramètres.

CHAPITRE 2

Modèle de l'apprenant et découverte des prérequis

Sommaire

2.1	Introduction	13
2.2	Learner Model	13
2.3	Modèle de compétence	19
2.4	Découverte des prérequis	22
2.5	Conclusion	23

2.1 Introduction

Le terme "apprentissage" fait souvent référence à l'apprentissage électronique ou e-learning. L'apprentissage en ligne, l'apprentissage à distance et l'apprentissage en ligne. L'apprentissage en ligne est connu comme des activités d'étude avec le soutien d'un ordinateur et d'un réseau (Fröschl, 2005, p. 12).

Dans l'apprentissage en ligne, la communication entre les apprenants et les enseignants ou les apprenants et les apprenants se fait sur le réseau ou sur Internet, les supports d'apprentissage sont souvent des documents électroniques tels que des pages Web et des fichiers au lieu de livres traditionnels sur papier [8]. Dans ce chapitre, nous décrivons la modélisation de l'apprenant compte tenu de son importance dans tous les domaines de recherche. Nous commençons par une présentation des notions du modèle de l'apprenant et des caractéristiques des méthodes utilisées pour la modélisation de l'apprenant. Puis nous terminons avec la Découverte des prérequis.

2.2 Learner Model

2.2.1 Définition

Winkels décrit le modèle de l'apprenant comme une structure de données qui reflète l'état des connaissances supposées de l'apprenant sur un domaine cible (le domaine d'apprentissage) [9]. Selon Self, le modèle de l'apprenant devrait inclure les informations suivantes : les connaissances et les idées fausses de l'apprenant, ses compétences et son comportement. Pour déterminer ce que l'apprenant a fait, Self souligne qu'il est nécessaire d'avoir un historique des interactions de l'apprenant avec le système [10].

Greer considère le modèle de l'apprenant comme une représentation abstraite des croyances, des connaissances et des compétences de l'apprenant dans le système, y compris l'historique des actions de l'apprenant qui peut être analysé et interprété [11].

2.2.2 Caractéristiques des méthodes utilisées pour la modélisation de l'apprenant

Il existe plusieurs modèles existants classés selon la nature des informations extraites par le modèle et les méthodes utilisées pour les traiter 2.1 et 2.2. [12]

Méthodes	Applications	Avantage	Limites
Modèle de superposition	Modélisation des connaissances de l'apprenant à travers le modèle de connaissance.	-Haute expressivité des problèmes complexes et flexibilité du raisonnement humain.	-Impossible de détecter les erreurs de l'apprenant
Modèle différentiel	- Identification des lacunes dans les connaissances des apprenants	- utilisation de la structure du modèle de connaissance.	-La complexité dépend de la structure du domaine. - Défaut de modéliser les erreurs de l'apprenant.
Modèle d'erreur	Identification des erreurs et fausses croyances de l'apprenant sur un domaine.	Identification des erreurs et fausses croyances de l'apprenant dans un domaine.	Difficulté à modéliser et à définir les erreurs.
Stéréotypes	Classification des apprenants selon des caractéristiques communes, les plus utilisées sont : le style d'apprentissage, le style cognitif, le niveau de connaissance et les préférences.	Initialisation de nouvelles caractéristiques de l'apprenant.	
Les ontologies	-Modélisation du modèle de connaissance. -Réutilisation et partage des ressources.	-Ajout de sémantique aux relations entre concepts. -Réutilisation des ressources.	
Logique floue	-Expression linguistique du niveau de connaissance. -Évaluation des apprenants. -Identification du style de l'apprenant.	-Expression du degré d'incertitude. -Grande expressivité linguistique et logique.	-Consommation de ressources -Pas d'apprentissage.

TABLE 2.1 : Caractéristiques des méthodes utilisées pour la modélisation des apprenants partie[1]

Méthodes	Applications	Avantage	Limites
Réseau Bayésien	<ul style="list-style-type: none"> -Apprendre et prédire le comportement et le raisonnement de l'apprenant. -Classement des apprenants. -Prédiction des performances des apprenants. 	<ul style="list-style-type: none"> -Apprendre et s'adapter à un environnement incertain. -Résoudre le problème « onlearning » -Fonction avec des données manquantes. 	<ul style="list-style-type: none"> -Consommation de ressources -Difficulté à comprendre les inférences (interprétation) -Grand nombre de variables utilisées.
Programmation génétique	<ul style="list-style-type: none"> - Recommandation d'un chemin de navigation optimal. -Prédiction des performances des apprenants. -Amélioration des algorithmes de classification. 	<ul style="list-style-type: none"> -Recherche globale -représentation flexible aux changements. -des recherches puissantes dans un espace de problèmes complexes et mal définis. 	<ul style="list-style-type: none"> -Temps d'exécution -Convergence vers les optima locaux
Réseau flou de neurones	<ul style="list-style-type: none"> -Évaluation et prédiction des performances des apprenants. 	<ul style="list-style-type: none"> -Grand Expressivement -Apprentissage -gestion des incertitudes. 	<ul style="list-style-type: none"> La complexité augmente avec le nombre d'entrées du réseau.

TABLE 2.2 : Caractéristiques des méthodes utilisées pour la modélisation des apprenants partie[2]

2.2.3 Composantes du modèle de l'apprenant

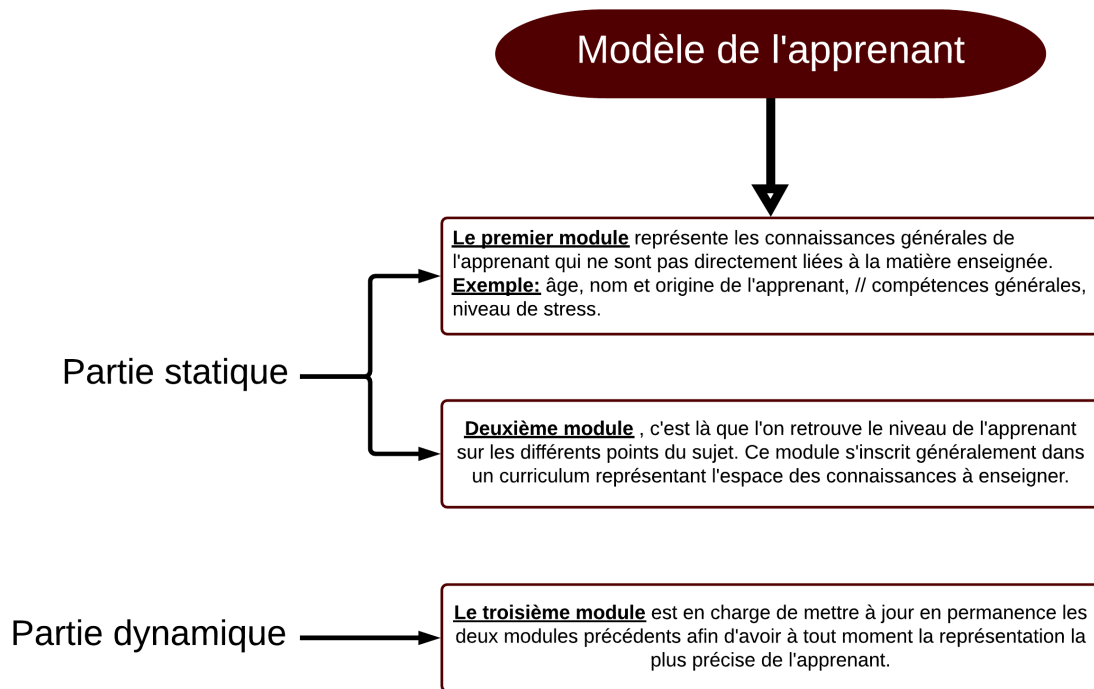


FIGURE 2.1 : Composantes du modèle de l'apprenant

Carr et Goldstein citent, par exemple, quatre informations nécessaires pour maintenir ce modèle : la difficulté du sujet, les questions directement posées par l'apprenant, la performance de l'apprenant et son expérience d'apprentissage [13]. Self définit le modèle de l'apprenant comme un 4-tuplet contenant les variables P (connaissance procédurale), C (connaissance conceptuelle), T (caractéristiques individuelles) et H (histoire) [14].

Un modèle de l'apprenant 2.1 comprend toujours des connaissances liées au domaine d'enseignement : ce que l'apprenant sait et ce que l'apprenant peut faire. Dans son état le plus complet, ce modèle comprend également des connaissances indépendantes du domaine enseigné. Certaines de ces connaissances sont liées à ses mécanismes d'apprentissage : comment fonctionne l'apprenant ? Comment découvre-t-il de nouveaux concepts ? De nouvelles techniques ? Etc. Les autres concernent les stratégies pédagogiques correspondantes : quels sont les types et modalités d'intervention les plus efficaces ? Etc.

Selon Nkambou, le modèle de l'apprenant est identifié en trois parties : [15]

2.2.3.1 Modèle cognitif

Description de l'état des connaissances de l'apprenant par rapport au sujet considéré par le bailleur de fonds. Ces informations concernent :

- **Capacités** : les informations sur les capacités reflètent le niveau de connaissances de l'apprenant. Robert M. Gagne a classé les capacités de l'apprenant en cinq catégories : l'information verbale, les compétences intellectuelles, les attitudes, la motricité et les stratégies cognitives.
- **Objectifs** : les informations sur les objectifs indiquent si l'apprenant a déjà atteint ou non un objectif.
- **Ressources** : les informations sur les ressources (exercices, problèmes, tests, etc.) reflètent le fait qu'une ressource a déjà été utilisée par un apprenant et le contexte dans lequel cette ressource a été utilisée.
- **Relations éventuelles** : les informations sur les relations indiquent si l'apprenant a réussi ou échoué à établir une relation (par exemple, analogie, abstraction, cas particulier, etc.) entre deux connaissances (et par conséquent, la connaissance d'une relation entre deux connaissances est aussi connaissance (méta-connaissance)).

2.2.3.2 Modèle d'inférence

Cette partie est une sorte de moteur d'inférence qui fonctionne en permanence pour ajuster le modèle de l'apprenant. Il contient des règles qui lui permettent de raisonner sur le modèle cognitif et sur le modèle psychologique (modèle affectif) pour inférer de nouvelles connaissances dans le modèle de l'apprenant.

2.2.3.3 Modèle émotionnel

Ce modèle est un ensemble de données qui nous permet d'identifier la personnalité et les différentes facettes d'un apprenant. Il contient des connaissances sur les caractéristiques permanentes ou momentanées particulières de l'apprenant. Parmi ceux-ci, nous avons :

- Connaissance des conditions mentales PUX. Par exemple, l'apprenant est spatial ou verbal, réfléchi ou impulsif, etc.
- Connaissance des sentiments et de la personnalité. Par exemple, l'apprenant est calme ou anxieux. Il est attentif ou distrait. Etc.

2.2.4 Utilitaire du modèle de l'apprenant

Selon Ragnemalm, il y a quatre utilisations du modèle de l'apprenant : [16]

- Importance du modèle pour la planification de l'éducation : quel contenu faut-il enseigner ?
- Présentation du contenu pédagogique : quelles expériences sont appropriées pour le contenu d'apprentissage ?
- Le feedback du système doit prendre en compte les connaissances précédemment mobilisées par l'apprenant, ainsi que le contexte d'apprentissage actuel.
- Traiter les idées fausses : en les signalant à l'apprenant, en fournissant un contre-exemple ou en suscitant une discussion

2.2.5 Types de modèles d'apprentissage

- **Implicit** : lorsque des informations décrivant le comportement de l'apprenant et influençant le cours de l'interaction avec le système sont incorporées dans le système.
- **Explicit** : lorsque les informations sur l'apprenant sont intégrées et codées dans le système de manière explicite pour gérer l'interaction avec l'apprenant.
- **Static** : lorsque les connaissances de l'apprenant sont déterminées avant toute utilisation et ne peuvent pas être modifiées en cours de session.
- **dynamic** : lorsque des données peuvent être ajoutées ou modifiées pendant la session.
- **Specific** : quand il peut être adapté à une catégorie d'apprenants.
- **Surface** : lorsqu'elle contient des informations limitées qui ne peuvent expliquer l'état cognitif de l'apprenant.
- **Deep** : lorsqu'il contient des informations plus représentatives de l'état cognitif de l'apprenant. [17]

2.2.6 Contenu du modèle de l'apprenant

Le modèle de l'apprenant représente ce que le système « sait » de l'apprenant. Ces informations peuvent être de nature cognitive, comportementale ou psychologique. Les premiers modèles se sont concentrés sur les aspects cognitifs et ont mis l'accent sur les connaissances

déclaratives, procédurales et heuristiques. Plus récemment, des modèles ont émergé pour représenter des aspects psychologiques : émotions et motivations. La modélisation de l'apprenant peut concerner un ou plusieurs aspects de l'apprenant : concepts, règles ou procédures de résolution maîtrisés, idées fausses, rapidité de résolution de problèmes, motivation à apprendre, capacité à réfléchir sur les connaissances apprises, aspects métacognitifs, etc. Le choix du contenu dépendra essentiellement du domaine d'enseignement, des objectifs didactiques et pédagogiques du système, des types d'interactions possibles avec l'apprenant, etc [18].

2.3 Modèle de compétence

Un modèle de compétences auquel il est fait référence ici ne concerne que les couches cachées du graphe des « couches de modélisation de l'apprenant » [19].

2.3.1 Granularité

La hiérarchie de granularité est une représentation courante d'un modèle d'apprenant [20]. Il décrit comment un domaine est décomposé en composants. Les composants de connaissance dans un modèle de domaine sont généralement décrits à différents niveaux de granularité. Une hiérarchie de granularité capture différents niveaux de détail dans un type de réseau sémantique. Les relations d'agrégation sont utilisées pour décrire les relations entre les composants de connaissances à différents niveaux. Les relations d'agrégation peuvent être utilisées pour diviser un composant de connaissance composite en plusieurs composants de connaissance à une granulométrie plus fine. Les observateurs sont généralement liés aux éléments de connaissance à un niveau plus fin. Les informations observées sont propagées par agrégation des liens vers les composants de connaissances aux niveaux plus grossiers. Le schéma de clustering AND-OR est proposé par Collins et al pour capturer les relations d'agrégation et les groupes équivalents dans leur hiérarchie de granularité [21].

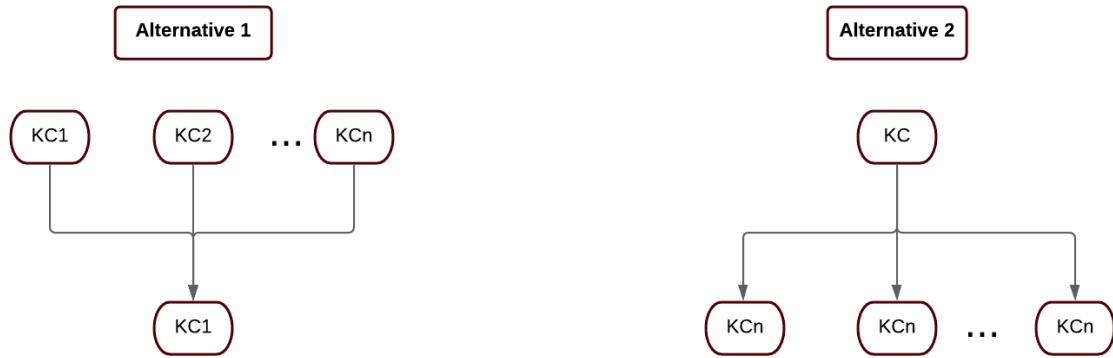


FIGURE 2.2 : Deux alternatives pour modéliser les relations d'agrégation.

Tchétagni et Nkambou ont proposé d'évaluer les connaissances des apprenants en logique propositionnelle à plusieurs niveaux de granularité. Ils ont utilisé le modèle alternatif 2 de la figure 2.2 pour représenter les relations d'agrégation dans leur hiérarchie. Ils ont souligné que dans cette architecture, il existe des restrictions sur la façon dont les preuves se propagent dans tout le réseau. Cela est dû au fait que deux nœuds enfants peuvent influencer leur parent, sans s'influencer mutuellement : ils sont séparés [22].

Carmona et Conejo ont utilisé dans la figure 2.2 le modèle alternatif 1 pour représenter les relations d'agrégation dans leur modèle d'apprenant utilisé dans MEDEA, un système ouvert pour le développement de systèmes de tuteurs intelligents. Certaines approches récentes ont abordé la granularité du modèle de compétences d'un point de vue statistique.

Dans ces approches, les modèles de compétences n'impliquent que les éléments de connaissance les plus fins, qui expliquent directement les comportements des apprenants. Un problème permanent dans un modèle d'apprenant est de savoir à quel niveau de granularité les compétences des apprenants doivent être modélisées [23]. Pardos et Heffernan ont exploré des modèles avec différents niveaux de granularité (modèles de compétences 1, 5, 39 et 106) et mesuré la précision de ces modèles en prédisant la performance des apprenants dans leur SIT, c'est-à-dire ASSISTment, ainsi que dans un test. Leurs résultats ont montré que plus la granularité du modèle de compétence est fine, meilleure est la prédiction des performances de l'apprenant [24].

2.3.2 Relations pré-requises

Des relations préalables existent généralement entre les éléments de connaissance de certains domaines.

Reye a analysé comment utiliser les réseaux bayésiens pour modéliser les relations anté-

rieures. Ils ont déclaré que les probabilités conditionnelles dans un réseau bayésien devraient remplir certaines conditions. Par exemple, si la composante de connaissance A est une condition préalable de la composante de connaissance B, l'équation II.1 doit être satisfaite. Cependant, ils ont également déclaré que la relation préalable n'est pas toujours stricte, de sorte qu'ils permettent une incertitude pour les probabilités conditionnelles. Les valeurs d'incertitude pour ces probabilités conditionnelles sont spécifiées par des experts dans leur méthode [25].

$$\begin{aligned} P(\text{learnerKnows}(A) \vee \text{learnerKnows}(B)) &= 1 \\ P(\text{learnerKnows}(B) \vee \text{learnerKnows}(A)) &= 0 \end{aligned} \quad (2.1)$$

Carmona et al. ont présenté les relations préalables à un modèle générique d'apprentissage du BN pour MEDEA, afin d'améliorer l'efficacité des mécanismes d'adaptation et le processus d'inférence. Ils ont utilisé une porte ET bruyante modifiée ou une porte OU bruyante modifiée pour modéliser les relations de prédicat [26].

Ferguson et al ont utilisé l'algorithme EM pour apprendre les paramètres cachés dans les BN et ont comparé le modèle de compétence plat (les compétences sont mutuellement indépendantes) avec le modèle de compétence hiérarchique (relations a priori entre des compétences données a priori) selon le critère d'information bayésien (BIC). Leurs résultats montrent qu'un modèle hiérarchique correspond mieux à leurs données que le modèle plat. [20==> a chercher]

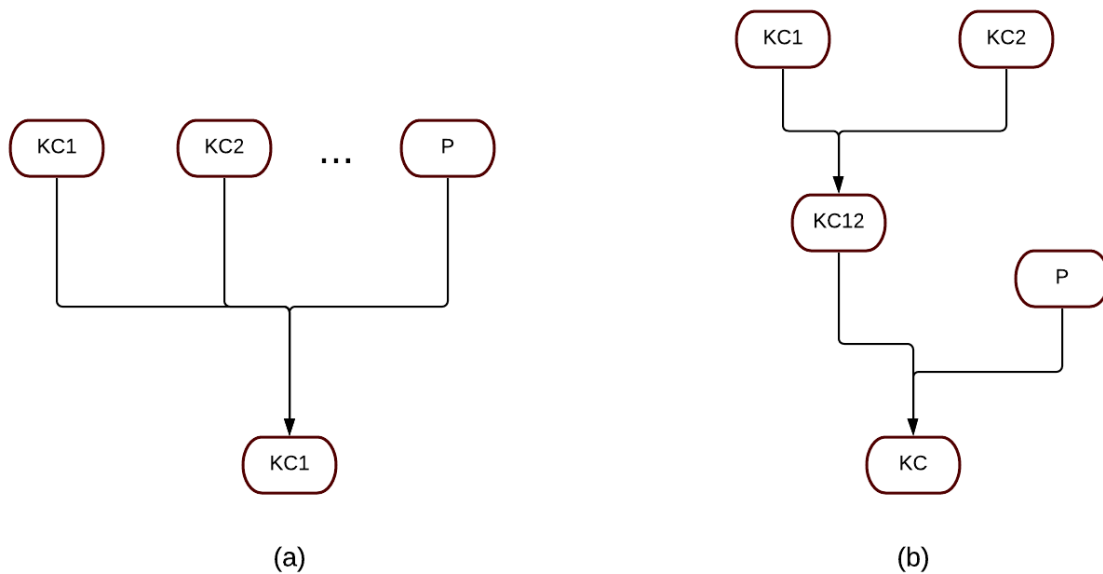


FIGURE 2.3 : Aggrégation de modélisation de réseau bayésien et relations préalables simultanément

Millan et coll. discuté d'un problème courant dans la modélisation des apprenants, à savoir modéliser simultanément les relations de pré-requis et de granularité. Si les deux sont inclus dans le même modèle, les relations avec des interprétations différentes sont mélangées et il est alors difficile de construire et de comprendre le modèle. Par exemple, si une compétence KC composite est composée de deux sous-compétences, KC1 et KC2, il existe également une compétence P qui est un prérequis pour KC. Les probabilités conditionnelles de K données à ses parents sont difficiles à préciser (figure 1.3 (a)). Ils ont proposé une solution qui consiste à regrouper des variables de même type en introduisant des variables intermédiaires (figure 1.3 (b)) [20].

2.4 Découverte des prérequis

Nous modélisons les compétences comme des variables continues qui représentent le degré auquel un apprenant a maîtrisé ou est conscient d'une compétence particulière. Nous traitons les éléments comme des variables continues qui reflètent le degré auquel un apprenant a correctement terminé une tâche. En pratique, la mesure de l'achèvement des tâches est souvent une variable binaire avec des valeurs = correct / incorrect. Un élément binaire peut cependant être considéré comme une projection d'un élément continu, et les corrélations entre éléments continus idéalisés peuvent être estimées en calculant la matrice de corrélation tétrachorique parmi les éléments binaires mesurés [27].

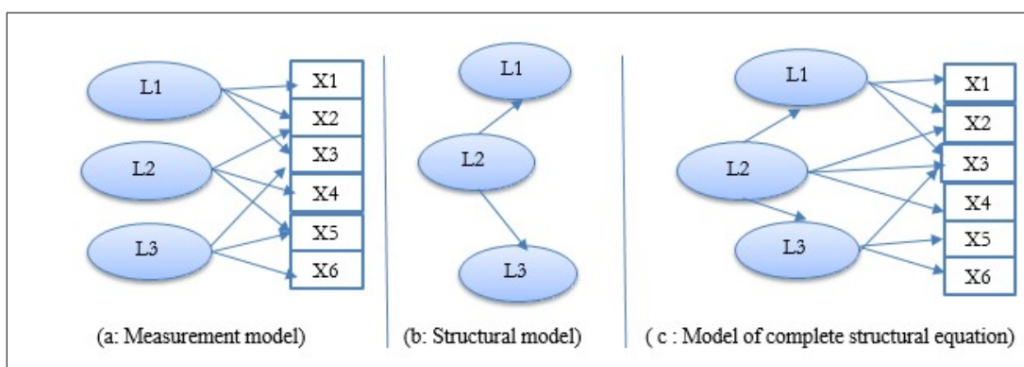


FIGURE 2.4 : Modèles d'équations structurelles

L'utilisation de la matrice Q consiste généralement à définir quels éléments « chargent » sur quelles compétences latentes. Nous pouvons définir un « modèle de mesure » qui relie les compétences latentes aux items mesurés 2.4. En modélisant les relations entre les compétences comme un modèle analytique causal du chemin entre les variables latentes 2.4, appelé

« modèle structurel », nous pouvons combiner le « modèle de mesure » et le « modèle structurel » pour former un modèle d'équation structurelle linéaire complet [2.4](#) [28].

En supposant que le modèle de mesure est connu, nous pouvons rechercher le modèle structurel avec l'algorithme de découverte causale PC, dans lequel les entrées sont les relations d'indépendance et d'indépendance conditionnelle qui figurent parmi les variables latentes. Nous calculons ou testons les relations d'indépendance entre les latents en construisant un modèle structurel séparé et en l'ajustant aux données pour chaque test d'indépendance particulière requis. Notre méthode de construction de modèle produit un test cohérent et éprouvé de chaque relation d'indépendance conditionnelle [27].

2.5 Conclusion

Tout au long de ce chapitre, nous avons présenté les principaux concepts liés au modèle de l'apprenant dans l'environnement de l'apprentissage basé sur le Web, qui permet aux apprenants de travailler en collaboration sur un projet ou de coconstruire le sens des concepts en utilisant la technologie comme un tuyau pour le développement des connaissances partagées, aussi, la modélisation des compétences est importante pour découvrir les prérequis d'un apprenant en représentant le degré auquel il a maîtrisé. Le prochain chapitre sera consacré à la présentation de l'exploration de données éducatives.

CHAPITRE 3

L’approche items-to-skills mapping

Sommaire

3.1	Introduction	25
3.2	Conclusion	25

3.1 Introduction

3.2 Conclusion

CHAPITRE 4

Fiabilité et l'accord pour les données d'observations

Sommaire

4.1	Introduction	27
4.2	Conclusion	27

4.1 Introduction

4.2 Conclusion

CHAPITRE 5

Inférence bayésienne et la théorie de la réponse aux items

Sommaire

5.1	Introduction	29
5.2	Conclusion	29

5.1 Introduction

5.2 Conclusion

CHAPITRE 6

Clustering

Sommaire

6.1	Introduction	31
6.2	Conclusion	31

6.1 Introduction

6.2 Conclusion

Deuxième partie

Contributions

CHAPITRE 7

Contribution and Results

Sommaire

7.1	Introduction	34
7.2	Conclusion	34

7.1 Introduction

7.2 Conclusion

Conclusion Générale et Perspectives

Conclusion Générale

Bibliographie

- [1] O. Scheuer and B. M. McLaren, *Educational Data Mining*, pp. 1075–1079. Boston, MA : Springer US, 2012.
- [2] J. Han and M. Kamber, *Data Mining : Concepts and Techniques*. 01 2000.
- [3] C. Romero and S. Ventura, “Educational data mining : A survey from 1995 to 2005,” *Expert Systems with Applications*, vol. 33, pp. 135–146, 07 2007.
- [4] C. Buche, *Un système tutoriel intelligent et adaptatif pour l'apprentissage de compétences en environnement virtuel de formation*. Theses, Université de Bretagne occidentale - Brest, Nov. 2005.
- [5] C. Romero, S. Ventura, M. Pechenizkiy, and R. Baker, *Handbook of Educational Data Mining*. 10 2010.
- [6] R. Baker and K. Yacef, “The state of educational data mining in 2009 : A review and future visions,” *Journal of Educational Data Mining*, vol. 1, pp. 3–17, 01 2009.
- [7] S. Bendjebar, *Utilisation des Techniques de Data Mining pour la Modélisation des Tuteurs*. Theses, Université 8 mai 1945 Guelma, 2016.
- [8] L. Nguyen and P. Do, “Learner model in adaptive learning,” *Proceedings of World Academy of Science, Engineering and Technology*, vol. 45, pp. 396–401, 01 2008.
- [9] R. Winkels, “User modelling in help systems.,” vol. 438, pp. 184–193, 06 1990.
- [10] A. Paiva and J. Self, “Tagus — a user and learner modeling workbench,” *User Modeling and User-Adapted Interaction*, vol. 4, pp. 197–226, 09 1994.
- [11] S. Greer, “Psycho-oncology : Its aims, achievements and future tasks,” *Psycho-Oncology*, vol. 3, pp. 87 – 101, 07 1994.
- [12] P. Brusilovsky, “Student model centered architecture for intelligent learning environments,” 01 1994.
- [13] I. Goldstein and B. Roberts, “Nudge, a knowledge-based scheduling program,” *Proceedings of IJCAI-77*, 01 1977.
- [14] J. Elster, *The Multiple Self*. Cambridge University Press, 1987.
- [15] R. Nkambou, “Modélisation des connaissances de la matière dans un système tutoriel intelligent : modèles, outils et applications /,” 06 2021.
- [16] E. Ragnemalm, “Collaborative dialogue with a learning companion as a source of information on student reasoning.,” pp. 650–658, 01 1996.

- [17] P. Holt, S. Dubs, M. Jones, and J. Greer, "The state of student modelling" in student modelling : The key to individualized knowledge-based inst," 06 2021.
- [18] D. Fragne, "Proposition de l'architecture de l'agent gestionnaire du modèle de l'apprenant dans un système tuteur multi-agents en apprentissage de la lecture : contribution au projet amical," 12 2009.
- [19] J. Gobert, M. Pedro, J. Raziuddin, and R. Baker, "From log files to assessment metrics : Measuring students' science inquiry skills using educational data mining," *Journal of the Learning Sciences*, vol. 22, pp. 521–563, 10 2013.
- [20] E. Millán, T. Loboda, and J. L. P. de-la Cruz, "Bayesian networks for student model engineering," vol. 55, no. 4, pp. 1663–1683, 2010.
- [21] J. Collins, J. Greer, and S. Huang, "Adaptive assessment using granularity hierarchies and bayesian nets," 02 1970.
- [22] J. M. P. Tchétagni and R. Nkambou, "Hierarchical representation and evaluation of the student in an intelligent tutoring system," in *Intelligent Tutoring Systems* (S. A. Cerri, G. Gouardères, and F. Paraguaçu, eds.), (Berlin, Heidelberg), pp. 708–717, Springer Berlin Heidelberg, 2002.
- [23] C. Carmona and R. Conejo, "A learner model in a distributed environment," in *Adaptive Hypermedia and Adaptive Web-Based Systems* (P. M. E. De Bra and W. Nejdl, eds.), (Berlin, Heidelberg), pp. 353–359, Springer Berlin Heidelberg, 2004.
- [24] Z. A. Pardos, N. T. Heffernan, B. Anderson, and C. L. Heffernan, "The effect of model granularity on student performance prediction using bayesian networks," in *User Modeling 2007* (C. Conati, K. McCoy, and G. Paliouras, eds.), (Berlin, Heidelberg), pp. 435–439, Springer Berlin Heidelberg, 2007.
- [25] J. Reye, "Student modelling based on belief networks," *International Journal of Artificial Intelligence in Education*, vol. 14, pp. 63–96, 2004. 1.
- [26] C. Carmona, E. Millán, J. L. Pérez-de-la Cruz, M. Trella, and R. Conejo, "Introducing prerequisite relations in a multi-layered bayesian student model," in *User Modeling 2005* (L. Ardissono, P. Brna, and A. Mitrovic, eds.), (Berlin, Heidelberg), pp. 347–356, Springer Berlin Heidelberg, 2005.
- [27] I. Goldin, r. scheines, and e. silver, "Discovering prerequisite relationships among knowledge components," 01 2014.
- [28] K. Bollen, "Structural equations with latent variables, new york," *Wiley Interscience*, 01 1989.