LOG6308 — Systèmes de recommandation

Modèles des compétences

Michel C. Desmarais

Génie informatique et génie logiciel
École Polytechnique de Montréal

Automne, 2017
(version 29 août 2017)

Modèles des compétences

ÉCOLE POLYTECHNIQUE MONTRÉAL

Types de modèles des compétences

Intérêts : on le retrouve plus ou moins implicitement dans les modèles de filtres d'informations. Les préférences, les votes, les comportements constituent, en fait, un modèle utilisateur.

Connaissances : très important pour les systèmes d'aide, les guides d'apprentissage et les systèmes tuteurs.

Buts : dans une certaine mesure, les intentions et les buts recherchés par l'utilisateur constituent aussi un modèle utilisateur, mais il est très dynamique. Il peut varier d'un instant à l'autre. On le verra avec les travaux sur la reconnaissance de plans.

Nous nous concentrerons ici sur les modèles de la connaissance.

Utilités

Les modèles de la connaissance d'un utilisateur peuvent servir à plusieurs applications :

- **Guide d'étude :** indique les matières sur lesquelles concentrer les efforts en vue d'un examen ou dans le cadre d'un cours.
- **Tuteur :** fournit un encadrement intelligent de matériel et exercices adaptés au réussites/échecs d'un étudiant.
- **Exerciseur :** adapte la difficulté des exercices/items au niveau de l'apprenant.

ÉCOLE POLYTECHNIQUE MONTRÉAL

Applications — guide d'étude, tuteurs, etc.

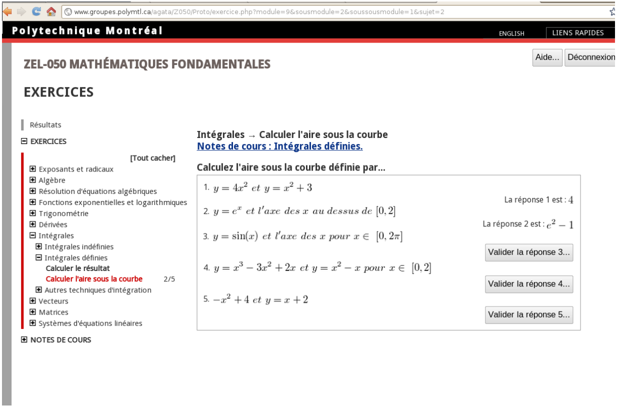
- *http://www.aleks.com/*
  Assistant pour l'étude des mathématiques.
- *http://quod.lib.umich.edu/j/jep/3336451.0006.110?rgn=main;view=fulltext*
  Tutoriel intelligent pour l'apprentissage de la physique.
- *http://www.brainbench.com/*
  Système pour la gestion des ressources humaines.
- *http://www.groupes.polymtl.ca/agata/Z050/Proto/*

Projet Agata Z050

Domaine	Évaluation
Algèbre et fonctions	A
Trigonométrie	A
Géométrie	C
Vecteurs et matrices	C
Calcul différentiel	A
Calcul intégral	A
Résultat global	**A**

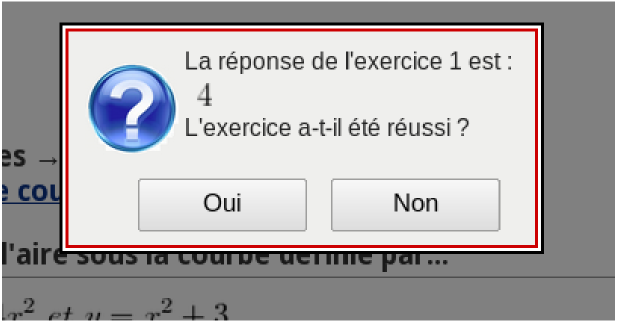
ÉCOLE POLYTECHNIQUE MONTRÉAL

Projet Agata Z050



ÉCOLE POLYTECHNIQUE MONTRÉAL

Projet Agata Z050



Projet Agata Z050

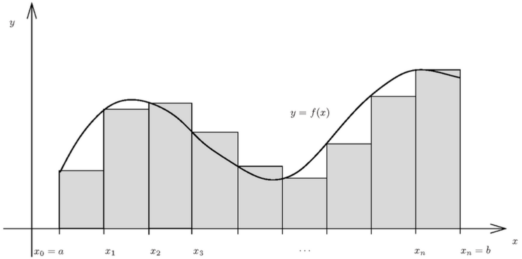
Exercices

donc $2I = \sec x \tan x + \ln|\sec(x) + \tan x| + C$

$I = \frac{1}{2}(\sec x \tan x + \ln|\sec(x) + \tan x| + C)$

9.2 Intégrale définie

Soit f une fonction définie sur un intervalle $[a, b]$.



- Subdivisons $[a, b]$ en n sous-intervalles $[x_i, x_{i+1}]$ de longueur égale $\Delta x = x_{i+1} - x_i$.
- Pour chaque sous-intervalle, soit $f(x_i)$ la valeur de la fonction à l'extrémité gauche.
- Soit $S_n = \sum_{i=0}^{n-1} f(x_i)\,\Delta x$ la **somme de Riemann** construite à partir de ces données.
- Soit $S = \lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{i=0}^{n-1} f(x_i)\,\Delta x$.

ÉCOLE POLYTECHNIQUE MONTRÉAL

Modèles de la connaissance

- Le modèle *stéréotype*
  L'utilisateur est classifié selon un *stéréotype*, une classe comme *expert*, *novice*, etc. Ce modèle a l'avantage de la simplicité et peut facilement être entraîné avec des méthodes probabilistes.
- Le modèle *hiérarchique*
  Les connaissances (concepts) sont organisées sous la forme d'une hiérarchie. La connaissance d'un noeud dépend de la connaissance de ses enfants.
- Le modèle *par item*
  Seuls des items *observables* (des questions ou exercices) sont utilisés pour modéliser la connaissance. De ces items, on peut déduire la connaissances de concepts non observables.

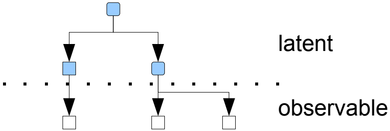
Autres dimensions des modèles de la connaissance

On peut aussi distinguer les modèles selon qu'ils comportent des paramètres latents ou observables et qu'ils sont construits par l'ingénierie de la connaissance ou par l'apprentissage automatique.

- Habiletés *latentes ou observables*
  Les concepts sont des habiletés latentes car ils ne sont jamais directement observés mais plutôt induits à partir d'indices observables, comme les réponses à des questions (items de tests).
- Estimation des paramètres par *apprentissage par ingénierie*
  L'estimation des paramètres (ex. probabilités conditionnelles) par apprentissage avec des données empiriques permet l'automatisation et évite la subjectivité d'un modèle dont les paramètres sont déterminés par un **expert du domaine**.

ÉCOLE POLYTECHNIQUE MONTRÉAL

Concepts latents et items observables



Les concepts sont normalement des paramètres latents puisqu'ils ne sont jamais. Dans la représentation graphique ci-dessus, les noeuds latents sont des concepts dans une structure hiérarchique dont les noeuds terminaux (feuilles de la hiérarchie) sont les items observables (question/items).

Modèles des compétences

ÉCOLE POLYTECHNIQUE MONTRÉAL

Note par sujet ?

Résultats regroupés par compétence/sujet :

Domaine	Évaluation
Algèbre et fonctions	A
Trigonométrie	A
Géométrie	C
Vecteurs et matrices	C
Calcul différentiel	A
Calcul intégral	A
Résultat global	**A**

Mais comment s'assurer quels items se rapportent à quels compétences/sujets

Résultats à un examen par sujet

	Résultats	Sujet
Question 1	1	X
Question 2	0	X
Question 3	1	Y
Question 4	1	Z
Question 5	1	Z
Question 6	0	Z

Score par sujet

Sujet X : 1/2 — Sujet Y : 1/1 — Sujet Z : 2/3

ÉCOLE POLYTECHNIQUE MONTRÉAL

Résultats à un examen par sujet

	Resultats	Sujet X	Sujet Y	Sujet Z
Question 1	1	1	0	0
Question 2	0	1	0	0
Question 3	1	0	1	0
Question 4	1	0	0	1
Question 5	1	0	0	1
Question 6	0	0	0	1

Q-matrice

Score par sujet

Sujet X : 1/2 — Sujet Y : 1/1 — Sujet Z : 2/3

Résultats à un examen par sujet

	Resultats	Sujet X	Sujet Y	Sujet Z
Question 1	1	1	0	0
Question 2	0	1	0	0
Question 3	1	0	1	0
Question 4	1	0	0	1
Question 5	1	0	0	1
Question 6	0	0	0	1

Q-matrice

Score par sujet

Sujet X : 1/2 — Sujet Y : 1/1 — Sujet Z : 2/3 $\Rightarrow$

$[\frac{1}{2}\,\frac{1}{1}\,\frac{2}{3}]^T = \mathbf{Q}^T[101110]^T$

Résultats à un examen par sujet

	Resultats	Sujet X	Sujet Y	Sujet Z
Question 1	1	1	0	0
Question 2	0	1	0	0
Question 3	1	0	1	0
Question 4	1	0	0	1
Question 5	1	0	0	1
Question 6	0	0	0	1

Q-matrice

Score par sujet

Sujet X : 1/2 — Sujet Y : 1/1 — Sujet Z : 2/3 $\Rightarrow$

$[\frac{1}{2}\frac{1}{1}\frac{2}{3}]^T = \mathbf{Q}^T[101110]^T$

(1) $\mathbf{S} = \mathbf{Q}^T \cdot \mathbf{R}$

(Les compétencese, **S**, sont le produit scalaire de **Q** et des résultats matrix **R**—si on normalise)

(2) $\mathbf{R} = \mathbf{Q} \cdot \mathbf{S}$

ÉCOLE POLYTECHNIQUE MONTRÉAL

Notions de Q-matrice

	Sujet X	Sujet Y	Sujet Z
Question 1	1	0	0
Question 2	1	0	0
Question 3	0	1	0
Question 4	0	0	1
Question 5	0	0	1
Question 6	0	0	1

Q-matrice

Variantes des Q-matrices :

- Peut avoir *une à plusieurs* compétences par item
- Modèle *conjonctif* toutes les compétences sont nécessaires au succès
- Modèle *disjonctif* : une seule compétence amène au succès
- Modèle *compensatoire* : chaque compétence contribue linéairement aux chances de succès

ÉCOLE POLYTECHNIQUE MONTRÉAL

Utilisation de Q-matrices en modélisation de l'apprenant

- Le sujet est très actif en recherche et soulève plusieurs questions :
  - Le modèle linéaire est-il approprié ?
  - Un modèle conjonctif est-il plus performant ?
  - Les prédictions sont-elles meilleures ou non aux autres modèles ?
- La factorisation peut apporter une meilleure performance prédictive, mais pas nécessairement se prêter à l'interprétation.

ÉCOLE POLYTECHNIQUE MONTRÉAL

Modèles des compétences

ÉCOLE POLYTECHNIQUE MONTRÉAL

Modèle bayésien simple

On utilise des éléments observables comme évidence pour l'appartenance à une classe donnée.

Posons trois items, X_1, X_2 et X_3, pouvant être soit *réussi* ($X = 1$) ou *échoué* ($X = 0$), et posons deux *stéréotypes*, *maître* ou *apprenti*, que nous exprimons comme étant respectivement $\theta = 1$ ou $\theta = 0$. Alors, le modèle bayésien consiste à évaluer :

$$P(\theta = 1 \mid X_1 = x_1, X_2 = x_2, X_3 = x_3)$$

où x_i représente le succès ou l'échec à l'item correspondant.

On abrège généralement l'expression précédente par :

$$P(\theta \mid X_1, X_2, X_3)$$

Exemple numérique d'estimation

Posons trois items, $\{X_1, X_2, X_3\}$, et une habiletée θ. Les items, comme l'habiletée, sont binomiales : elles peuvent prendre deux valeurs, $\{0, 1\}$. Le tableau suivant donne la distribution des fréquences de 200 individus pour chacune des combinaisons X_i.

j	X_1	X_2	X_3	$f(\theta_j = 1)$	$f(\theta_j = 0)$	$P(\theta_j)$
1	0	0	0	0	10	0,08
2	0	0	1	2	10	0,21
3	0	1	0	8	25	0,26
4	0	1	1	10	15	0,41
5	1	0	0	25	13	0,65
6	1	0	1	12	17	0,42
7	1	1	0	8	6	0,56
8	1	1	1	35	4	0,88

La probabilité $P(\theta_j)$ est estimée par $\frac{f(\theta_j=1)+1}{(\sum_v f(\theta_j=v))+2}$

Problème de combinatoire et croissanse exponentielle de la demande de données

- *Estimation des probabilité conditionnelles* Dans le tableau qui précède, on peut directement estimer $P(\theta_3)$, par exemple :

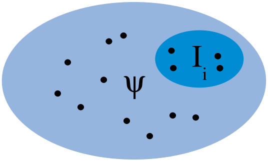
$$P(\theta_3) = P(\theta | X_1 = 0, X_2 = 1, X_3 = 0) = \frac{8+1}{25+8+2} = 0,26$$

- *Explosion combinatoire* Tout va bien car nous avons trois items et 200 cas de calibration. Pour 10 items, le tableau aurait 2^{10} lignes (1024 entrées à calibrer) et il nous faudrait des dizaines de milliers de cas pour obtenir un estimé de toutes les combinaisons. Il devient donc rapidement impossible de calibrer les probabilités conditionnnelles.

Modèles des compétences

ÉCOLE POLYTECHNIQUE MONTRÉAL

Modèles de sous-ensemble (*overlay* et espaces de connaissances



- Le domaine de connaissances est représenté par un ensemble ψ.
- L'ensemble contient des items de connaissances, $\{a_1, a_2, \ldots, a_n\}$
  $\psi \equiv \{a_1, a_2, \ldots, a_n\}$

- Ces items sont directement **observables**
- L'état de la connaissance d'un individu, I_i, est représenté par un sous ensemble de ψ

$$I_i \subseteq \psi$$

- **Granularité** potentielle de 2^n

Notions de fermeture

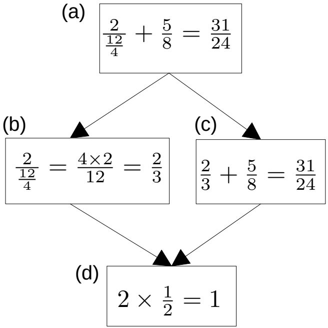
Supposons que l'on représente l'état connaissance d'un individu par un sous-ensemble d'unités de connaissances dans un ensemble qui, lui, représente le domaine, ψ. Ces unités peuvent être évalués par une ou plusieurs questions ou items.

S'il est vrai que :

- la mise en commun (*union*) de l'état de connaissance et de compétences de deux individus est aussi l'état (hypothétique) d'un autre individu, et que
- l'état de leur connaissance commune (*intersection*) est aussi l'état (hypothétique) d'un individu,

alors il s'en suit que les états de connaissances, les sous-ensembles possibles de ψ, peuvent être représentés par un ordre partiel, c.-à-d. un graphe dirigé.

Un exemple



Exemple d'un espace de connaissance composé de 5 items ($\psi = \{a, b, c, d\}$) et formant une ordre partiel qui contraint les états de connaissances à $\{\emptyset, \{d\}, \{b, d\}, \{c, d\}, \{b, c, d\}, \{d, b, c, a\}\}$.

Inférence et choix de la question

Une fois le réseau construit, il s'agit de choisir la prochaine question et d'effectuer un nouvel estimé des probabilité de succès à chaque question. Voyons en premier lieu l'inférence (estimé de probabilité d'un item) avec un réseau donné et une série d'observations.

Les inférences se font selon le modèle de "combinaison des évidences" (voir le rappel de la théorie de Bayes) :

$$O(H \mid E_1, E_2, \ldots, E_n) = O(H) \prod_i^n L\bullet_{E_i H} \qquad \text{où}$$

$L\bullet_{E_i H}$ représente $LS_{E_i H}$ ou $LN_{E_i H}$ selon le cas.

Dans notre cas, $O(H)$ représente les chances de succès à l'item que l'on veut estimer, et les évidences, $E_1, E_2, \ldots, E_n$, représente les observations d'items réussis ou échoués. Il s'agit donc de calculer leur LS et LN respectifs.

Le choix de la question

Dans le cas où l'on a le choix de la prochaine question à poser, différentes options existent.

1. Poser la question qui optimisera l'évaluation de la connaissance et réduira l'incertitude
   - Réduction de l'entropie
   - Calcul de l'information de Fisher
2. Adapter la question au niveau de connaissance de l'individu

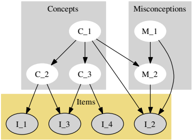
En pratique, la première option aura dans une certaine mesure l'effet d'adapter la question au niveau d'expertise de l'utilisateur, mais dans certains cas sans être optimal car on vise un taux de succès plus élevé pour ne pas décourager l'utilisateur (ex. avec des enfants).

Modèles des compétences

ÉCOLE POLYTECHNIQUE MONTRÉAL

Modèles alternatifs

- *Modèles de réseaux de bayes*
  On peut aussi modéliser les liens entre concepts eux-mêmes et entre les concepts et les items par différentes simplifications comme les "noisy OR-gates" ou tout autre structure probabiliste qui a une signification.

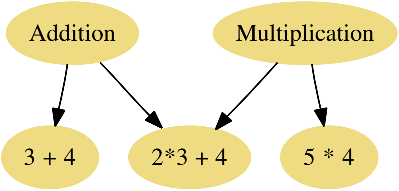


- *Modèles à base de règles*
  Un autre modèle utilise des règles pour à établir des diagnostics précis et analyser des comportements plus complexes. Il peut être combiné à un modèle probabiliste.
- *Modèles simple*
  Finalement, on peut aussi utiliser une matrice et une simple formule de sommation pour indiquer le niveau de maîtrise de différents concepts, à l'instar d'un examen.

Les modèles graphiques, bayésiens

- Très répandus dans le domaine de la modélisation des connaissances des apprenants.

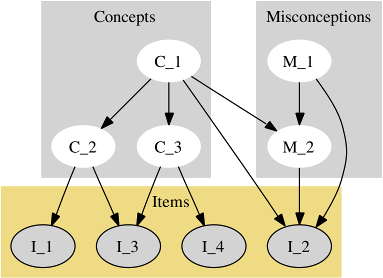
- Les **Réseaux bayésiens** sont particulièrement utilisés pour modéliser les concepts entre eux.
- Exemple d'un réseau bayésien comportant des questions (items), des concepts et des concepts erronés (*misconceptions*)



Les modèles graphiques, bayésiens

- Très répandus dans le domaine de la modélisation des connaissances des apprenants.

- Les **Réseaux bayésiens** sont particulièrement utilisés pour modéliser les concepts entre eux.
- Exemple d'un réseau bayésien comportant des questions (items), des concepts et des concepts erronés (*misconceptions*)

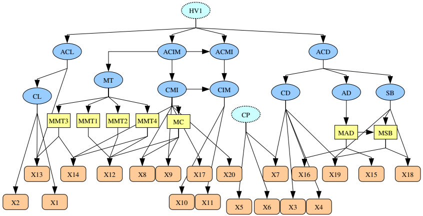


Les modèles graphiques, bayésiens

Un réseau dans le domaine de l'arithmétique, Vomlel (2004)

- 20 questions
- 20 concepts et concepts erronés
- 149 répondants au questionnaire

Réseau bayésien en arithméthique des fractions

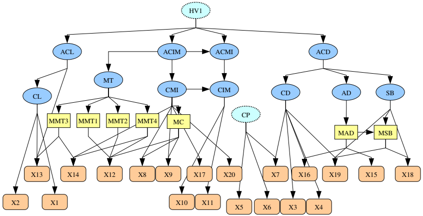


Les modèles graphiques, bayésiens

Un réseau dans le domaine de l'arithmétique, Vomlel (2004)

- Partiellement *appris* avec des données indépendantes de la maîtrise des concepts
- Partiellement *construit à la main*
- Inférence par des techniques répandues en réseaux bayésiens (*junction tree*)

Réseau bayésien en arithméthique des fractions



L'inférence avec un réseau bayésien

1. Certaines structures comme les arbres se prête à un calcul exact
2. Des structures libres doivent recourir à des algorithmes d'optimisation, comme par exemple l'algorithme Expectation-Maximisation (EM).
3. Cet algorithme peut prendre plusieurs formes, dont voici les grandes lignes d'une variante :
   1. En commençant par les racines, on assigne des valeurs aléatoires mais selon les probabilités initiales
   2. On propage les assignations vers les noeuds enfants en respectant les valeurs assignées aux noeuds parents
   3. On créé ainsi plusieurs échantillons à partir desquels on recalcule les probabilités conditionnelles à partir des assignations aléatoires
   4. Les nouvelles probabilités sont réutilisées pour créer de nouveaux échantillons
4. Souvent on utilise un modèle *ad hoc* qui simplifie les calculs mais ne respecte pas la sémantique des inférences bayésiennes

Modèles des compétences

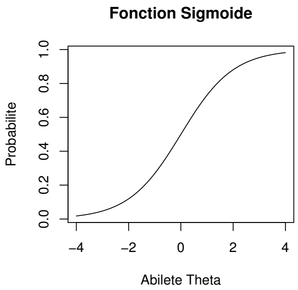
Un exemple

Présumons que le succès d'un individu à répondre correctement à une série de questions, $(X_1, X_2, \ldots, X_n)$, dépend de son habileté, θ. La probabilité d'observer une série spécifique de réponses, $(x_1, x_2, \ldots, x_n)$ est donc :

$$P(x_1, x_2, \cdots, x_n|\theta) \propto \prod_i P(x_i|\theta)$$

Le problème consiste donc à trouver la valeur de θ qui maximisera cette probabilité. Il nous faut un modèle de $P(X_i|\theta)$ comme celui de la figure ci-contre, qui correspond à la *fonction sigmoïde* :

$$P(X_i|\theta) = \frac{1}{1 + e^{-\theta}}$$



Un exemple (suite)

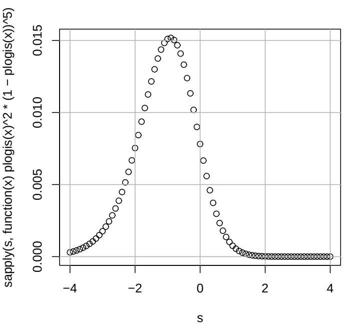
Si notre individu a répondu à 7 questions et obtenu :

$$(0, 1, 1, 0, 0, 0, 0)$$

alors la probabilité d'observer cette séquence s'il avait une habileté $\theta = 1$ serait proportionnelle à

$$0,731^2 \times (1 - 0,731)^5 = 0,001$$

car $P(X_i = 1 | \theta = 1) = \frac{1}{1+e^{-1}} = 0,731$.
La probabilité de (0,1,1,0,0,0,0) pour $\theta = -1$ correspond à $0,015$



Le graphique ci-dessus indique la probabilité pour les différentes valeurs de θ. La valeur maximale se situe près de $\theta = -1$

ÉCOLE POLYTECHNIQUE MONTRÉAL

Le modèle TRI (*IRT*)

Le modèle le plus répandu pour modéliser les réponses à des items d'un test est celui de la Théorie des Réponses aux Items (TRI, ou *Item Response Theory—IRT*). Ce modèle s'inspire de la fonction sigmoïde mais il comprend deux paramètres supplémentaires :

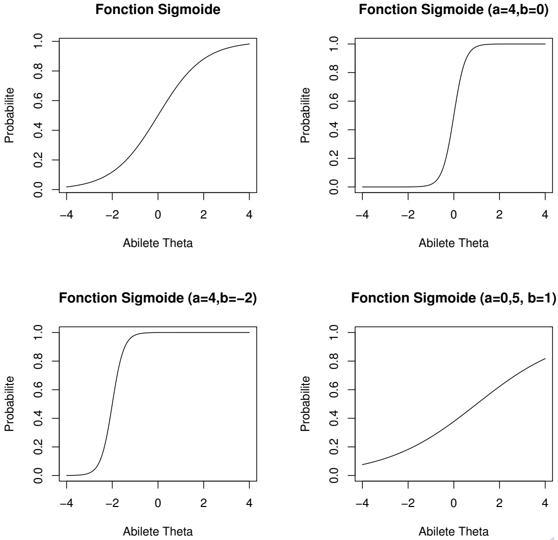
a : la *discrimination* d'un item qui détermine la pente ;

b : la *difficulté* d'un item qui détermine l'abscisse ;

La fonction résultante est celle de la réponse caractéristique à un item (*Item Characteristic Curve—ICC*) :

$$P(X_i|\theta) = \frac{1}{1 + e^{-a_i(\theta - b_i)}}$$

Différentes valeurs de a et b



Modèle TRI (suite)

L'implication d'inclure deux paramètres est que chaque item possède alors son propre niveau de difficulté et sa capacité de discriminer entre $\theta = 1$ et $\theta = 0$.

Ces paramètres peuvent être estimés selon la méthode de la vraisemblance maximale : on choisi leur valeurs pour maximiser la probabilité d'observer un échantillon donné de réponses.

ÉCOLE POLYTECHNIQUE MONTRÉAL