MADA-ETI, ISSN 2220-0673, Vol.2, 2016, www.madarevues.gov.mg

APPLICATION DE LA THEORIE DES CHAINES DE MARKOV DANS LE CALCUL DU COEFFICIENT D'EFFICACITE DU SYSTEME EDUCATIF

RAZAFINDRABE Raymond[1-] et RASTEFANO Elisée[2]

Laboratoire de Recherche Systèmes Embarqués, Instrumentations et Modélisation des Systèmes et Dispositifs Electroniques

Ecole Doctorale en Sciences et Techniques de l'Ingénierie et de l'Innovation

Université d'Antananarivo

[1]razafraym@yahoo.fr. - [2]rastefano_el@yahoo.fr

Résumé

Le Coefficient d'Efficacité, un indicateur synthétique du système éducatif est habituellement dégagé par comptage du nombre d'années-élèves du modèle de diagramme de flux. Sa valeur dépend du nombre de redoublement fixé pour l'élaboration du diagramme. Le modèle confectionné à partir de la théorie des chaînes de Markov présente une expression directement calculable à partir des taux de flux; Il suppose, néanmoins une hypothèse, trop théorique, de nombre infini de redoublements autorisés. En examinant, cependant, le fonctionnement du modèle de diagramme des flux, un ajustement des paramètres de calcul s'avère nécessaire, et il se trouve que l'expression obtenue du modèle de chaîne de Markov est tout à fait applicable même en se référant au modèle de diagramme de flux.

Summary

The educational synthetic indicator, Coefficient of Efficiency is usually obtained by counting the number of pupil-years in a flow diagram model. Its value depends on the number of authorized repetition. Model based on Markov chain theory shows an expression directly calculable from flow rates; nevertheless, it presupposes an infinite number of authorized repetitions. By examining the flow diagram functioning, parameter for calculation can be adjusted and finally, formula from Markov chain theory is applicable even on referring to a flow diagram model.

Introduction

La théorie des chaînes de Markov peut servir au traitement de données dans l'analyse d'un système susceptible de se trouver dans différents états. Le présent article concerne un aspect du système éducatif, portant sur le processus de parcours scolaire et consacré plus précisément au calcul d'un indicateur d'appréciation de l'efficacité, appelé coefficient d'efficacité. Il présente en premier lieu le modèle diagramme de flux habituellement utilisée dans le secteur éducatif pour le calcul de tel indicateur. En second lieu sera dégagé une formulation théorique basée sur un processus markovien. Enfin, le troisième volet fera constater l'alignement des résultats des deux modèles après ajustement des paramètres de la méthode de diagramme de flux.

1. Modèle de Diagramme de Flux

Le coefficient d'efficacité est un indicateur synthétique de mesure de l'efficacité du système éducatif au niveau d'un cycle donné [1]. Il s'agit du ratio rapportant le nombre minimum d'année-élèves[1] requis pour la formation des sortants (diplômés) du cycle au nombre d'années-élèves réellement consommés. La situation idéale est ainsi celle où tous les sortants terminent leur parcours scolaire sans redoublement. La situation réelle consomme davantage d'années-élèves en raison des redoublements, mais aussi des abandons scolaires.

Le calcul du coefficient d'efficacité se ramène essentiellement à la détermination du nombre d'années-élèves consommés. Dans la pratique, cette quantité est dégagée à partir d'un diagramme de flux reconstituant le parcours scolaire d'une cohorte d'élèves. Le modèle représente les flux d'élèves suivant deux axes représentant d'une part l'année d'étude, et d'autre part l'année scolaire.

A la fin de l'année scolaire, l'effectif $E_{i,j}$---où le premier indice représente l'année d'études et le second le nombre de redoublement subi par la proportion d'élèves en question--- se décompose en trois parties en fonction des taux de promotion (p_i) de redoublement (r_i) et d'abandon (a_i)

- les promus : p_i x $E_{i,j}$
- les redoublants : r_i x $E_{i,j}$
- les cas d'abandon : a_i x $E_{i,j}$

[1] Une année scolaire passée dans une classe (ou année d'études) par un élève est comptée comme une année-élève

L'élaboration du diagramme utilise, pour le calcul des flux, les taux calculés à partir des statistiques scolaires de deux années successives. Il est supposé que :

- les taux de flux (p_i, r_i et a_i) se rapportant à chacune des années d'études restent inchangés pour les différentes années scolaires;
- le nombre maximum de redoublement autorisé fixé pour l'ensemble du cycle; les redoublements se transforment en abandon lorsque le seuil est atteint; et
- la représentation ne concerne qu'une cohorte : aucun élève ne s'ajoute au cours des différentes années scolaires.

Pour une meilleure lisibilité du diagramme, l'effectif $E_{1,0}$ de la cohorte est généralement fixé à 1000 ; telle initialisation n'entraine pas d'impact sur le calcul du coefficient d'efficacité qui s'exprime sous forme ratio. La Figure 1 suivante donne le diagramme de flux d'une cohorte reconstituée pour le cas d'un cycle comportant *n* années d'études et d'un nombre de redoublement plafonné à *k*.

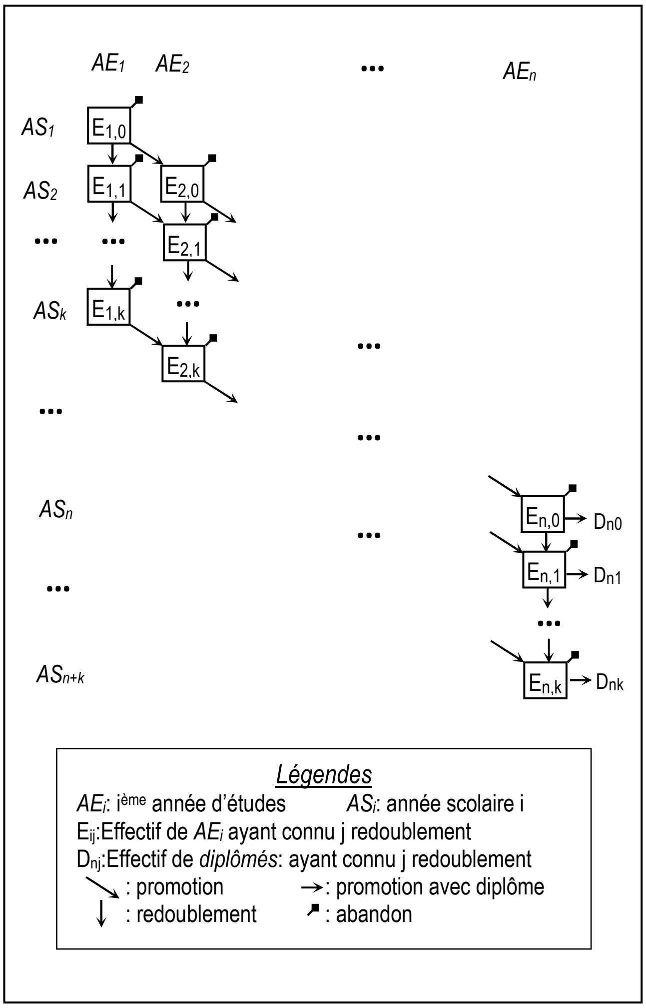

Figure 1. Diagramme de Flux d'élèves

Les nombre d'années-élèves idéal et nombre d'années-élèves réellement consommés peuvent être comptés directement à partir du diagramme:

$$CE = \frac{n\sum_{j=1}^{k} D_{n,j}}{\sum_{i=1}^{n}\sum_{j=1}^{k} E_{i,j}}$$

Il convient de remarquer en particulier que l'hypothèse sur le nombre de redoublement a une influence déterminante sur la valeur de l'expression.

2. Modélisation par la Chaîne de Markov

Comme présenté plus haut, les élèves d'une classe se répartissent dans des destinations différentes suivant les taux de flux p_i, r_i et a_i. Ces derniers peuvent être considérés comme des probabilités de passer d'un état à un autre.

En considérant comme hypothèse que les taux de flux restent constants dans le temps (exprimé dans le présent cas en termes d'années scolaire) le parcours scolaire peut être modélisé comme un processus markovien illustré par la Figure 2 [2].

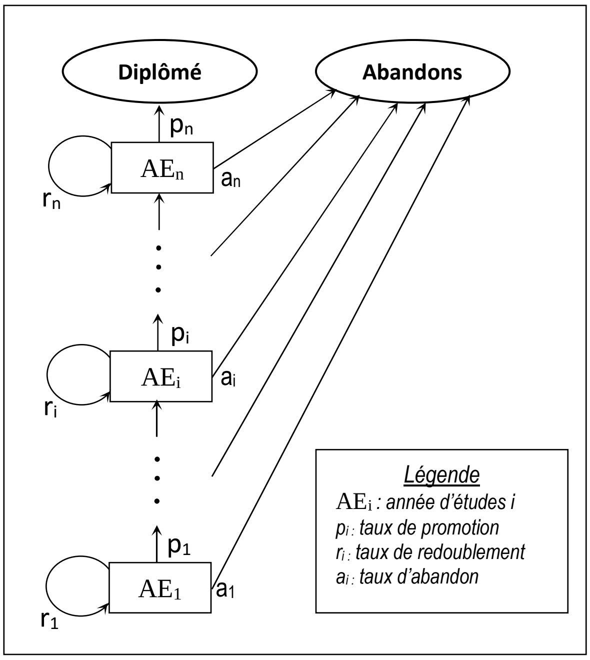

Figure 2. Diagramme de Transition du Parcours Scolaire

Il s'agit plus précisément d'une chaîne de Markov absorbante, ayant deux états absorbants: l'état « diplômé » et l'état «abandon». En effet, arrivé à l'un de ces deux états, l'élève ne pourrait plus quitter pour d'autre état. La matrice de transition (de l'état figuré en ligne à l'état figuré en colonne) correspondant à ce diagramme peut être mise sous la forme canonique suivante, les lignes et les colonnes indiquent les états (A = abandon, D=diplômé, AE_i=année d'études) :

A D AE_n … AE_2 AE_1

A
D
AE_n
…
AE_2
AE_1

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & \cdots \\ a_n & p_n & r_n & 0 & \cdots & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & 0 & 0 & 0 \\ a_2 & 0 & \cdots & p_2 & r_2 & 0 & 0 \\ a_1 & 0 & \cdots & \cdots & 0 & p_1 & r_1 \end{pmatrix}$$

Cette matrice est de la forme :

$$T = \begin{pmatrix} I & 0 \\ A & Q \end{pmatrix},$$ I étant la matrice identité 2 sur 2

La matrice fondamentale de la chaîne est $F = (I-Q)^{-1}$

$$\begin{array}{cc} & \begin{array}{cccc} \boxed{AE_n} & \dots & \boxed{AE_2} & \boxed{AE_1} \end{array} \\ F = & \begin{pmatrix} \frac{1}{(1-r_n)} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \frac{p_2 \dots p_{n-1}}{(1-r_2)\dots(1-r_n)} & \cdots & \frac{1}{(1-r_2)} & 0 \\ \frac{p_1 \dots p_{n-1}}{(1-r_1)\dots(1-r_n)} & \cdots & \frac{p_1}{(1-r_1)(1-r_2)} & \frac{1}{1-r_1} \end{pmatrix} \begin{array}{c} \boxed{AE_n} \\ \dots \\ \boxed{AE_2} \\ \boxed{AE_1} \end{array} \end{array}$$

Les éléments de cette matrice donnent le nombre moyen de passages de l'état figuré en ligne à l'état figuré en colonne [3]. Ainsi, la dernière ligne porte le nombre de passage dans les années d'études successives. Il s'agit, en termes d'analyse du système éducatif, du nombre d'années élèves consommées par un individu correspondant à chaque année d'études.

En considérant une cohorte comportant initialement $E_{1,0}$ élèves, et en posant $p_0 = 1$(dans le but de faciliter l'écriture, car en fait le taux de promotion p_0 n'existe pas), le nombre total d'années-élèves est alors :

$$NAE = E_{1,0} \sum_{j=1}^{n} \frac{\prod_{i=0}^{j-1} p_i}{\prod_{i=1}^{j}(1-r_i)} \quad \text{avec } p_0 = 1$$

Concernant l'effectif des diplômés, il est donné par Le nombre de passage à la dernière année d'études AE_n pondéré par le taux de réussite aux examens qui se présente comme étant le taux de promotion p_n de cette classe.

$$\sum_{j=0}^{k} D_{n,j} = E_{1,0} \frac{p_n \prod_{i=1}^{n-1} p_i}{\prod_{i=1}^{n}(1-r_i)} = \frac{\prod_{i=1}^{n} p_i}{\prod_{i=1}^{n}(1-r_i)}$$

Finalement, l'expression du coefficient d'efficacité est :

$$CE = \frac{\dfrac{n \prod_{i=1}^{n} p_i}{\prod_{i=1}^{n}(1-r_i)}}{\sum_{j=1}^{n} \dfrac{\prod_{i=0}^{j-1} p_i}{\prod_{i=1}^{j}(1-r_i)}}$$

avec $p_0 = 1$

Dans cette expression, le calcul de l'indicateur s'effectue immédiatement à partir des taux de flux relevés, ne nécessitant alors pas de comptage particulier de nombre d'années-élèves comme c'est le cas dans le modèle de diagramme de flux. Seulement, la validité de la formule est soumise à l'hypothèse de nombre théoriquement infini de possibilité de redoublement de classe; ce qui n'est pas conforme à la pratique.

3. De l'Amélioration du Modèle de Diagramme des Flux à la Généralisation des résultats de la Modélisation par la Théorie des Chaînes de Markov

En examinant de près le fonctionnement du modèle de diagramme des flux utilisés pour les calculs du nombre d'années-élèves consommées, des améliorations peuvent être apportées.

Il est à rappeler que les taux de flux utilisés dans l'élaboration du diagramme sont déterminés à partir des statistiques scolaires de deux années successives. Les statistiques utilisées ne distinguent point les composantes de l'effectif d'une classe en fonction du nombre de redoublement connu par l'élève. Les taux calculés constituent ainsi des taux moyens relevés pour l'ensemble des composantes ($E_{i,1}$, $E_{i,2}$,… $E_{i,k}$ dans le Schéma 1). Le diagramme des flux utilisent, cependant, à la fois ces taux moyens et d'autres taux si bien que l'ensemble ne correspond plus aux taux moyens relevé.

En prenant le cas de l'année d'études i :

- le taux de promotion utilisé pour le diagramme est p_i quelle que soit l'année-scolaire;
- le taux de redoublement est r_i pour $E_{i,0}$ à $E_{1,k-1}$ mais 0 pour $E_{1,k}$;
- le taux d'abandon est a_i pour $E_{i,0}$ à $E_{1,k-1}$ mais $a_i + r_i$ pour $E_{1,k}$.

Ainsi, en utilisant dans le diagramme le taux de redoublement observé r_1 pour l'année d'études AE_1, on obtient une représentation biaisée de la situation de la cohorte reconstituée car au lieu d'avoir comme taux de redoublement r_1, le taux observé, on a, relativement au diagramme, un taux moyen égal à :

$$r'_1 = \frac{(E_{1,0} + \cdots + E_{1,k-1})r_1}{E_{1,0} + \cdots + E_{1,k}}$$

Pour corriger cette distorsion, on devrait utiliser une autre valeur du taux de redoublement qui donne comme valeur moyenne la valeur observée r_1. Si ce taux est ρ_1, alors, il doit vérifier la relation :

$$r_1 = \frac{(E_{1,0} + \cdots + E_{1,k-1})\rho_1}{E_{1,0} + \cdots + E_{1,k}}$$

$$= \frac{E_{1,1} + \cdots + E_{1,k}}{E_{1,0} + \cdots + E_{1,k}}$$

$$= 1 - \frac{E_{1,0}}{E_{1,0} + \cdots + E_{1,k}}$$

On en déduit la somme des années-élèves correspondant à la première année d'études AE_1:

$$\sum_{j=0}^{k} E_{1,j} = \frac{1}{1 - r_1} E_{1,0}$$

Pour l'année d'études AE_2, en utilisant les taux corrigé ρ_1 et ρ_2, les différentes composantes du nombre d'années-élèves s'expriment respectivement:

Le taux de redoublement correspondant est nul pour $E_{2,k}$ et ρ_2 pour les autres composantes; donc

$$r_2 = \frac{\rho_2 \sum_{j=0}^{k-1} E_{2,j}}{\sum_{j=0}^{k} E_{2,j}}$$

Ou bien

$$\rho_2 = \frac{r_2 \sum_{j=0}^{k} E_{2,j}}{\sum_{j=0}^{k-1} E_{2,j}}$$

On a $E_{2,0} = E_{1,0}\, p_1$

$E_{2,1} = E_{2,0}\, \rho_2 + E_{1,1}\, p_1$

…

$E_{2,k} = E_{2,k-1}\, \rho_2 + E_{1,k}\, p_1$

D'où :

$$\sum_{j=0}^{k} E_{2,j} = \rho_2 \sum_{j=0}^{k-1} E_{2,j} + p_1 \sum_{j=1}^{k} E_{1,j}$$

$$= \frac{r_2 \sum_{j=0}^{k} E_{2,j}}{\sum_{j=0}^{k-1} E_{2,j}} \sum_{j=0}^{k-1} E_{2,j} + p_1 \sum_{j=1}^{k} E_{1,j}$$

$$= r_2 \sum_{j=0}^{k} E_{2,j} + p_1 \sum_{j=1}^{k} E_{1,j}$$

On en tire l'expression de la somme des années-élèves correspondant à la deuxième année d'études AE_2:

$$\sum_{j=0}^{k} E_{2,j} = \frac{p_1 \sum_{j=1}^{k} E_{1,j}}{1 - r_2} = \frac{p_1}{(1 - r_1)(1 - r_2)} E_{1,0}$$

Le raisonnement peut être continué par récurrence en supposant que pour l'année d'études AE_i, i>1 la somme des années-élèves est :

$$\sum_{j=1}^{k} E_{i,j} = \frac{\prod_{j=1}^{i-1} p_j}{\prod_{j=1}^{i} (1 - r_j)} E_{1,0}$$

Pour l'année d'études AE_{i+1}:

$$\rho_{i+1} = \frac{r_{i+1} \sum_{j=0}^{k} E_{i+1,j}}{\sum_{j=0}^{k-1} E_{i+1,j}}$$

Les effectifs de l'année d'études AE_{i+1} peuvent être décomposés en deux groupes qui sont

- les redoublants de l'année d'études AE_{i+1} s'élèvant à

$$\rho_{i+1} \sum_{j=0}^{k-1} E_{i+1,j} = \frac{r_{i+1} \sum_{j=0}^{k} E_{i+1,j}}{\sum_{j=0}^{k-1} E_{i+1,j}} \sum_{j=0}^{k-1} E_{i+1,j}$$

$$= r_{i+1} \sum_{j=0}^{k} E_{i+1,j}$$

- les promus de l'année d'études AE_i formulés par :

$$p_i \sum_{j=1}^{k} E_{i,j} = p_i \frac{\prod_{j=1}^{i-1} p_j}{\prod_{j=1}^{i} (1 - r_j)} E_{1,0} = \frac{\prod_{j=1}^{i} p_j}{\prod_{j=1}^{i} (1 - r_j)} E_{1,0}$$

Donc :

$$\sum_{j=0}^{k} E_{i+1,j} = r_{i+1} \sum_{j=0}^{k} E_{i+1,j} + \frac{\prod_{j=1}^{i} p_j}{\prod_{j=1}^{i} (1 - r_j)} E_{1,0}$$

$$\sum_{j=0}^{k} E_{i+1,j} = \frac{\prod_{j=1}^{i} p_j}{\prod_{j=1}^{i+1} (1 - r_j)} E_{1,0}$$

CQFD

On retrouve alors intégralement les expressions déduites du cas de nombre infini de redoublement telles que présentées dans la matrice fondamentale de la modélisation par la théorie des chaînes de Markov

En définitive,

L'expression du nombre d'années-élèves consommées dans le système éducatif ne dépend pas du nombre de redoublement autorisé mais uniquement des valeurs des taux de flux relevés :

$$NAE = E_{1,0} \sum_{j=1}^{n} \frac{\prod_{i=0}^{j-1} p_i}{\prod_{i=1}^{j} (1 - r_i)}$$

avec $p_0 = 1$ et $E_{1,0}$ l'effectif initial de la cohorte fictif

Telle expression peut être appliquée dans tout calcul faisant intervenir le nombre d'années-élèves consommés dont, en particulier, la détermination de l'indicateur synthétique Coefficient d'Efficacité qui est alors

$$CE = \frac{\dfrac{n \prod_{i=1}^{n} p_i}{\prod_{i=1}^{n} (1 - r_i)}}{\sum_{j=1}^{n} \dfrac{\prod_{0}^{j-1} p_i}{\prod_{1}^{j} (1 - r_i)}}$$

avec $p_0 = 1$

Conclusion

En théorie, les formules dérivées des chaînes de Markov, s'appliquent dans l'hypothèse une succession quasi-infinie de séquences d'événements. Dans le présent cas traitant de données sur le parcours scolaires, il se trouve que les résultats coïncident exactement avec ceux du modèle de diagramme des flux qui suppose un nombre limité de séquences, comme le suppose la réalité effective. Du coup, la procédure (relativement lourde) habituellement pratiquée dans le calcul d'indicateur comme le coefficient d'efficacité du système éducatif peut être substituée par une simple application de formule

REFERENCE

[1] Institut de Statistiques de l'UNESCO, Indicateurs de l'Education, Directive Techniques, 2009

[2] Kemeny John G., Laurie Snell J. *Finite Markov Chains, "Generalization of a Fundamental Matrix"*. New York: Springer-Verlag, 1975.

[3] Krewera Germain. Graphes, Chaîne de Markov et Quelques Applications Economiques. Paris: Dalloz 1972.