

Système de recommandation de cours basé sur les soft skills : Une approche utilisant les algorithmes génétiques

Luis Alberto Pinos Ullauri^{1,2,3}[0000–0002–3694–3487], Alexis Lebis¹, Abir Karami⁴, Mathieu Vermeulen¹ Anthony Fleury¹, and Wim Van Den Noortgate^{2,3}

¹ IMT Nord Europe, Centre for Digital Systems, Lille, France

² KU Leuven, Faculty of Psychological and Educational Sciences, Leuven, Belgium

³ imec ITEC KU Leuven, Leuven, Belgium

⁴ Université Catholique de Lille, Lille, France

Résumé L'importance croissante prise ces dernières années par les soft skills dans l'enseignement supérieur a encouragé les recherches sur leur relation avec les activités pédagogiques. Ce travail s'intéresse à l'incidence des soft skills sur la recommandation d'un ensemble de cours aux étudiants en fonction à la fois du profil de formation souhaité et des attentes du monde professionnel. Trois fonctions d'évaluation et deux méthodes de compensation sont proposées pour simuler différentes réalités socio-professionnelle et exprimer la différence entre la maîtrise des soft skills estimés heuristiquement et le score attendu à la fin du programme.

Keywords: Systèmes de recommandation · Soft Skills · Algorithmes Génétiques · Modèles Mixtes Linéaires · Fitness

Abstract. The increasing attention regarding soft skills in higher education has promoted research to study its relation with pedagogical activities. This work presents a framework to recommend a set of courses to students based on their soft skills considering their expected profile at the end of an academic program. Three score functions and two compensation methods are proposed to handle the difference between the heuristically estimated proficiency and the minimum expected proficiency of their soft skills by the end of the program.

Keywords: Recommender Systems · Soft Skills · Genetic Algorithms · linear mixed models · fitness

1 Introduction

L'intérêt pour les soft skills a été au départ porté principalement par l'industrie mais a maintenant une place importante dans les établissements d'enseignement supérieur (EES). La conceptualisation et la formalisation du terme «soft skills» font l'objet de débats entre les chercheurs. Cependant, une interprétation générale, telle que présentée par [1], les décrit comme un groupe de compétences non techniques, intrapersonnelles et interpersonnelles que nous pouvons utiliser

en diverses situations. De leur côté, [3] ont réalisé trois études différentes dans lesquelles ils ont constaté que la demande de soft skills variait en fonction des disciplines des étudiants, soutenant ainsi l'idée des profils soft skills différents, comme le soutient [2]. En plus, la notion de soft skills peut s'entremêler avec celle de compétences, néanmoins, les compétences peuvent également faire référence à des aptitudes techniques, alors que les soft skills sont flexibles et multidisciplinaires (*i.e.* gestion du stress, leadership). Récemment, [5] a proposé un système de recommandation (SR) pour individualiser le cursus des étudiants par rapport aux compétences et le cadre formel associé. Néanmoins, cette recherche fondamentale ne traite pas du caractère particulier des soft skills, ni de l'intérêt que peuvent avoir les étudiants concernant les cours proposés et l'effet associé. Notre problématique est alors de pouvoir permettre la recommandation de parcours ancrés dans ces singularités.

2 Contributions et Propositions

Un scénario hypothétique, basé sur une situation réelle dans un ESS en France, est présenté afin d'illustrer le problème. Les travaux de [5] ont prouvé que le problème de recommandation de parcours individualisé – où un étudiant dans un cursus doit suivre un nombre fini de cours dans des périodes temporels spécifiques comme des semestres – est au moins NP-Complet. Les auteurs proposent une modélisation générique : les cours sont piochés dans un catalogue de cours en autorisant les combinaisons avec répétition. Pour limiter la complexité du problème, nous avons décidé de nous placer dans un cas de SR où les combinaisons sont sans répétition. Dans notre scénario, les étudiants doivent suivre $n = 11$ cours parmi $m = 16$ (C_{16}^{11} .) options possibles et quatre spécialisations différentes sont envisagées. Nous supposons également que les cours ne sont pas spécifiques à une spécialisation et nous considérons pour le moment trois soft skills indépendants. Les soft skills sont traités comme des valeurs continues et pouvaient être mesurés avec IRT [4]. Chaque spécialisation définit des attentes (*i.e.* seuils) décrivant la maîtrise des soft skills attendue, montré dans le tableau 1.

TABLE 1 – Définition des seuils par spécialisation

Spécialisation	Soft Skill 1	Soft Skill 2	Soft Skill 3
Spéc. 1	0	0	0
Spéc. 2	0.5	0	-1
Spéc. 3	0	0.5	-1
Spéc. 4	-1	0	0.5

2.1 Génération de données

Nous proposons le modèle présenté dans l'équation 1 pour générer la maîtrise du soft skill i par l'étudiant s au semestre t . β_{i0} représente la maîtrise initiale moyenne sur le soft skill i d'un étudiant moyen. β_{ic} représente l'effet fixe du cours c sur le soft skill i . α_{ics} représente un effet aléatoire de l'étudiant s vers le cours c et le soft skill i . En outre, I_{cst} se rapporte à l'intérêt de l'étudiant s pour le choix du cours c au semestre t . De plus, θ_{is} représente l'effet aléatoire

Système de recommandation de cours basé sur les soft skills: Une approche utilisant les algorithmes génétiques

de l'étudiant s dans le développement du soft skill i . Et, enfin, ϵ_i est une erreur résiduelle du soft skill i .

$$SSkill_{ist} = \beta_{i0} + \sum_{c=1}^{16} ((\beta_{ic} + \alpha_{ics})D_{cst} + I_{cst}) + \theta_{is} + \epsilon_i \quad (1)$$

où D_{cst} est une variable fictive qui est égale à 1 si l'étudiant s a suivi le cours c au semestre t . Nous avons utilisé ce modèle pour produire des jeux de données simulés sur 1000 étudiants, du début jusqu'à bout d'un programme de 4 semestres. Les distributions de paramètres et les effets sont indiqués dans le tableau 2.

TABLE 2 – Distributions et valeurs des paramètres

Paramètre	Distribution	Effet
β_{i0}	-	(1.5,1,0.5)
α_{ics}	$N(\mu = 0, \sigma_{i=1} = 0.05, \sigma_{i=2} = 0.75, \sigma_{i=3} = 0.1)$	-
I_{cst}	$N(\mu = 0, \sigma = 0.1)$	-
θ_{is}	$N(\mu = 0, \sigma_{i=1} = 0.15, \sigma_{i=2} = 0.2, \sigma_{i=3} = 0.25)$	-
ϵ_i	$N(\mu = 0, \sigma_{i=1} = 0.2, \sigma_{i=2} = 0.15, \sigma_{i=3} = 0.1)$	-

2.2 Algorithmes génétiques et Fitness

Pour trouver des solutions à ce problème de recommandation basées sur les soft skills et l'intérêt de l'étudiant, nous avons utilisé des algorithmes génétiques (AG). L'évaluation de la solution est réalisée par une fonction multidimensionnelle mono-objectif sur la combinaison de cours de l'étudiant s et leur apport en termes de soft skills. Trois fonctions sont utilisées pour évaluer les combinaisons : 1) une différence linéaire entre le soft skill estimé et le seuil, 2) une fonction quadratique par morceaux à gauche et à droite des seuils et 3) une fonction par morceaux exponentielle (cf. Équation 2).

$$score_{is} = f(\widehat{SSkill}_i) = \begin{cases} -\epsilon^{1.5(Seuil_m - \widehat{SSkill}_i)} & \text{if } \widehat{SSkill}_i < Seuil_m \\ \epsilon^{\widehat{SSkill}_i - Seuil_m} & \text{if } \widehat{SSkill}_i \geq Seuil_m \end{cases} \quad (2)$$

3 Analyse et résultats préliminaires

Le paramétrage utilisé pour notre AG ($X=0.66$, $M=1/16$, $Nb_{executions} = 100$) a été empiriquement défini. Les jeux de données générés ont été divisés avec un ratio de 80% (entraînement)- 20% (test). Nos résultats se basent sur trois métriques : le pourcentage 1) des étudiants qui ont passé tous les seuils, 2) de ceux qui ont passé au moins un seuil mais pas tous et 3) d'aucun seuil validé. Le tableau 3 montre nos résultats préliminaires. L'on peut voir que toutes les fonctions proposées permettent d'augmenter (entre 0% et 11%) le pourcentage moyen d'étudiants qui réussissent les seuils dans les 4 spécialisations.

4 Discussion et conclusions

Notre AG a été implémenté en Python. Les expérimentations ont été conduites sur un processeur Intel^(R) Core^(TM) i7-10810U @ 1.10GHz et 32 Go de RAM.

TABLE 3 – Comparaisons métriques en moyenne entre les fonctions. Les trois pourcentages sont ordonnés de la manière suivante : (tout, partiellement et aucun réussi)

Fitness	Spéc. 1	Spéc. 2	Spéc. 3	Spéc. 4
Sans Rec.	(29%,39%,32%)	(26%,60%,14%)	(25%,62%,13%)	(24%,63%,13%)
Comp. L	(34%,34%,32%)	(30%,48%,22%)	(33%,46%,21%)	(35%,50%,15%)
Par. Comp. L	(31%,33%,36%)	(28%,52%,20%)	(28%,50%,22%)	(32%,53%,15%)
Comp. QR	(33%,35%,32%)	(33%,46%,21%)	(32%,49%,19%)	(37%,47%,16%)
Par. Comp. QR	(28%,33%,39%)	(26%,49%,25%)	(26%,51%,23%)	(29%,55%,16%)
Comp. E	(32%,34%,34%)	(30%,49%,20%)	(32%,48%,20%)	(35%,50%,15%)
Par. Comp. E	(31%,32%,37%)	(26%,54%,20%)	(27%,49%,24%)	(29%,54%,17%)

Le temps d'exécution d'une recommandation était de l'ordre de 15 secondes (soit de 50 minutes par chaque 200 étudiants). Nous avons donc parallélisé les calculs pour réduire les temps totaux d'exécution. Les résultats préliminaires montrent une polarisation de la population, avec toutefois des changements positifs (étudiants ayant réussi tous les seuils) plus importants. Néanmoins, les métriques utilisées pour le moment ne peuvent à elles seules déterminer quelle fonction de score est à préférer pour ce type de problème, car chacune d'entre elles repose sur une logique différente. En plus, il y a certaines considérations à prendre en compte, qui constituent nos perspectives. Premièrement, la complexité croissante du problème avec le nombre de cours et sa réalité terrain. Deuxièmement, les fonctions de fitness proposées fonctionnent avec des mesures continues de soft skills, et devraient être adaptées dans le cas des mesures catégoriques. Finalement, nos résultats soulèvent des questions éthiques à étudier.

Remerciement

Cette contribution a été réalisée dans le cadre du projet de recherche SUCCESS, financé par l'i-SITE ULNE.

Références

1. Almonte, R. : A Practical Guide to Soft Skills : Communication, Psychology, and Ethics for Your Professional Life. Routledge, Taylor & Francis Group (2021), <https://doi.org/10.4324/9781003212942>
2. Barrie, S. : Understanding what we mean by generic attributes of graduates. Higher Education **51**, 215–241 (01 2006)
3. Chamorro-Premuzic, T., Artheche, A., Bremner, A., Greven, C., Furnham, A. : Soft skills in higher education : Importance and improvement ratings as a function of individual differences and academic performance. Educational Psychology **30**, 221–241 (03 2010)
4. Hartig, J., Höhler, J. : Multidimensional irt models for the assessment of competencies. Studies in Educational Evaluation **35**(2), 57–63 (2009). <https://doi.org/https://doi.org/10.1016/j.stueduc.2009.10.002>, assessment of Competencies
5. Lebis, A., Humeau, J., Fleury, A., Vermeulen, M. : Le cursus académique personnalisé dans une approche par compétences avec érosion : étude d'un nouveau problème fondamental. In : 10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain. pp. 94–105. Fribourg, Switzerland (2021)