

ÉLÉMENT DE PORTFOLIO 01



Publication

1 DÉFINITION DE CET ÉLÉMENT

Titre de l'élément : Harrak, F., Bouchet, F., & Luengo, V. (2019). From Student Questions to Student Profiles in a Blended Learning Environment. *Journal of Learning Analytics*, 6(1), 54-84.

URL de l'élément : <https://doi.org/10.18608/jla.2019.61.4>

2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

Cet article a été réalisé dans le cadre du travail de thèse de Fatima Harrak [2], co-encadrée par Vanda Luengo et François Bouchet, et par ailleurs récompensé pour une version française de cet article [4] par le prix Monique Grandbastien du meilleur article de la conférence EIAH 2019¹, la conférence internationale francophone principale dans notre domaine. Il s'agit en outre du journal phare de la communauté internationale de recherche autour des learning analytics, portée par l'association SOLAR², qui est un des thèmes de recherche de l'équipe.

En termes de contenu, il est particulièrement représentatif des efforts mis en place par l'équipe pour mélanger des approches symboliques et numériques puisque :

- ▶ il s'appuie sur un schéma de codage de questions d'étudiants constitué sur un cas d'étude particulier (des questions d'étudiants de médecine en situation de classe inversée),
- ▶ l'automatisation de l'annotation des questions a été menée en testant à la fois un système à base de règles et un ensemble de classifieurs entraînés sur des données annotées,
- ▶ un clustering est mené sur les questions d'élèves ainsi annotées.

Il correspond aussi à notre volonté de croiser différentes sources de données puisqu'ici les clusters constitués sur la base des questions sont ensuite expliqués à partir de variables liées à l'activité et à la performance en classe des étudiants. Enfin il montre le besoin fort de pluridisciplinarité dans notre domaine de recherche, puisqu'il s'appuie en partie pour la comparaison des questions posées vs. votées sur le cadre théorique Active-Constructive-Interactive proposé en psychologie de l'apprentissage par [1].

3 PRÉSENTATION DE CET ÉLÉMENT

Ce travail vise à montrer comment l'analyse automatique de questions posées par des étudiants peut permettre d'améliorer l'expérience d'apprentissage des étudiants et des enseignants. Nous nous sommes intéressés aux questions posées par des étudiants en première année d'études de médecine/pharmacie (PACES) sur une plateforme en ligne et qui est utilisée par les enseignants pour préparer des sessions de question-réponse en présentiel. L'objectif est double : côté enseignant, il s'agit de proposer une catégorisation des questions non seulement en fonction de leur nature mais aussi en fonction de ce que l'on sait des élèves qui ont posé des questions de même nature par le passé, afin de sélectionner au mieux les questions auxquelles des réponses sont à donner dans un contexte de concours où il est essentiel de ne pas favoriser certains profils d'étudiants. Côté élève, il s'agit de pouvoir leur proposer à terme un feedback sur la nature des questions qu'ils posent au cours de l'année pour éventuellement les aider à améliorer la pertinence de celles-ci.

Partant d'un corpus de 6457 questions, nous avons développé un schéma de codage multi-dimensionnel pour celles-ci, identifiant notamment la nature de la question (réexplication, approfondissement, vérification), la modalité d'explication (exemple, schéma, correction), le type d'explication demandé (la manière, la raison, le rôle, le lien entre concepts...).

1. Environnements Informatiques pour l'Apprentissage Humain – <https://eiah2019.sciencesconf.org/>

2. SOciety for Learning Analytics Research – <https://www.solaresearch.org/>

Nous avons ensuite exploré différentes stratégies pour automatiser l'annotation des questions : une approche à base de règles et des classifieurs automatiques à base d'apprentissage machine³. Ces annotateurs automatiques ont été utilisés pour annoter l'ensemble de notre corpus.

Grâce à ce corpus annoté, nous avons pu mener une étude exploratoire des liens existants entre les questions posées et les caractéristiques des étudiants, en regardant plus particulièrement :

- ▶ les étudiants posant des questions vs. ceux qui n'en posaient pas
- ▶ si la proportion des types de questions posées était liée de la performance de l'étudiant
- ▶ si l'analyse de la dynamique des questions posées au cours du temps était liée à la performance des étudiants
- ▶ si les caractéristiques de l'étudiant étaient associées à la nature des questions posées

Ce dernier point a été exploré en utilisant un clustering à base de K-Means pour constituer des clusters uniquement sur la base des proportions de questions posées à différents moments du semestre au sein de 4 cours différents. Nous avons ensuite caractérisé les clusters à l'aide d'attributs non utilisés pour le clustering comme les notes, la participation, le nombre de questions posées et la popularité de celles-ci (en termes de votes reçus). Nous avons pu mettre en avant que dans tous les cours considérés, deux clusters de nature similaire apparaissaient systématiquement (des étudiants en-dessous de la moyenne avec des questions populaires, et des étudiants au-dessus de la moyenne avec des questions peu populaires).

Enfin nous avons répliqué ces analyses sur les mêmes cours mais sur différentes années scolaires pour montrer la stabilité de ces résultats et qu'il était donc possible de prédire le profil d'un élève à partir de la nature des questions posées en ligne.

4 RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Michelene TH Chi. Active-constructive-interactive : A conceptual framework for differentiating learning activities. *Topics in cognitive science*, 1(1) :73–105, 2009.
- [2] Fatima Harrak. *Analyse de questions d'apprenants et de profils associés dans des environnements en ligne*. These de doctorat, Sorbonne université, October 2019.
- [3] Fatima Harrak, François Bouchet, and Vanda Luengo. Categorizing students' questions using an ensemble hybrid approach. In Collin F. Lynch, Agathe Merceron, Michel Desmarais, and Roger Nkambou, editors, *Proc. of the 12th International Conference on Educational Data Mining*, pages 312–317, Montréal, Canada, July 2019. myPub116.
- [4] Fatima Harrak, François Bouchet, and Vanda Luengo. Comparaison de questions posées et votées en ligne dans le cadre d'une classe inversée. In Julien Broisin, Eric Sanchez, Amel Yessad, and Françoise Chenevotot, editors, *Actes de la 9ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain*, pages 121–132, Paris, France, June 2019.

3. Dans un travail ultérieur, nous avons montré qu'il était possible de les combiner avec une approche ensembliste par stacking pour combler leurs lacunes respectives et améliorer la performance globale [3]